

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification: A01N 37/18; 43/04, C12Q 1/00, 1/02, 1/68, C12N 5/00, 5/06, 15/00, 15/06, 15/09, 15/10, 15/11, G01N 33/53</p>	<p>A2</p>	<p>(43) International Publication Number: WO 98/48000 (43) International Publication Date: 10 December 1998 (10.12.98)</p>
<p>(21) International Application Number: PCT/US98/11422</p> <p>(22) International Filing Date: 4 June 1998 (04.06.98)</p> <p>(30) Priority Data: 60/048,915 6 June 1997 (06.06.97) US 60/048,882 6 June 1997 (06.06.97) US (Continued on the following page)</p> <p>(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 877 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US).</p>	<p>Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). FLORENCE, Charles [US/US]; (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FAN, Ping [CN/US]; Apartment 302, 335 West Side Drive, Gaithersburg, MD 20878 (US). WEI, Ying-Fei [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9720 Maggett Farm Drive, Potomac, MD 20854 (US). CARTER, [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US).</p> <p>(74) Agents: HOOVER, Kenley, Kenley Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.</p>	
<p>(54) Title: 207 HUMAN SECRETED PROTEINS</p>		

(Continued)

60/048,892	6 June 1997 (06.06.97)	US	60/057,651	5 September 1997 (05.09.97)	US
60/048,901	6 June 1997 (06.06.97)	US	60/057,769	5 September 1997 (05.09.97)	US
60/048,900	6 June 1997 (06.06.97)	US	60/057,643	5 September 1997 (05.09.97)	US
60/048,893	6 June 1997 (06.06.97)	US	60/057,645	5 September 1997 (05.09.97)	US
60/048,964	6 June 1997 (06.06.97)	US	60/057,668	5 September 1997 (05.09.97)	US
60/048,884	6 June 1997 (06.06.97)	US	60/057,635	5 September 1997 (05.09.97)	US
60/048,894	6 June 1997 (06.06.97)	US	60/057,627	5 September 1997 (05.09.97)	US
60/048,971	6 June 1997 (06.06.97)	US	60/057,667	5 September 1997 (05.09.97)	US
60/048,885	6 June 1997 (06.06.97)	US	60/057,666	5 September 1997 (05.09.97)	US
60/049,375	6 June 1997 (06.06.97)	US	60/057,764	5 September 1997 (05.09.97)	US
60/048,881	6 June 1997 (06.06.97)	US	60/057,644	5 September 1997 (05.09.97)	US
60/048,880	6 June 1997 (06.06.97)	US	60/057,765	5 September 1997 (05.09.97)	US
60/048,896	6 June 1997 (06.06.97)	US	60/057,762	5 September 1997 (05.09.97)	US
60/049,020	6 June 1997 (06.06.97)	US	60/057,775	5 September 1997 (05.09.97)	US
60/048,876	6 June 1997 (06.06.97)	US	60/057,634	5 September 1997 (05.09.97)	US
60/048,895	6 June 1997 (06.06.97)	US	60/057,777	5 September 1997 (05.09.97)	US
60/049,019	6 June 1997 (06.06.97)	US	60/057,628	5 September 1997 (05.09.97)	US
60/048,916	6 June 1997 (06.06.97)	US	60/057,776	5 September 1997 (05.09.97)	US
60/048,970	6 June 1997 (06.06.97)	US	60/057,760	5 September 1997 (05.09.97)	US
60/048,972	6 June 1997 (06.06.97)	US	60/057,761	5 September 1997 (05.09.97)	US
60/048,949	6 June 1997 (06.06.97)	US	60/057,771	5 September 1997 (05.09.97)	US
60/048,974	6 June 1997 (06.06.97)	US	60/057,770	5 September 1997 (05.09.97)	US
60/048,883	6 June 1997 (06.06.97)	US	60/057,649	5 September 1997 (05.09.97)	US
60/048,897	6 June 1997 (06.06.97)	US	60/057,774	5 September 1997 (05.09.97)	US
60/048,898	6 June 1997 (06.06.97)	US	60/057,648	5 September 1997 (05.09.97)	US
60/049,373	6 June 1997 (06.06.97)	US	60/057,642	5 September 1997 (05.09.97)	US
60/048,917	6 June 1997 (06.06.97)	US	60/057,629	5 September 1997 (05.09.97)	US
60/048,962	6 June 1997 (06.06.97)	US	60/057,778	5 September 1997 (05.09.97)	US
60/048,878	6 June 1997 (06.06.97)	US	60/057,763	5 September 1997 (05.09.97)	US
60/049,374	6 June 1997 (06.06.97)	US	60/057,584	5 September 1997 (05.09.97)	US
60/048,875	6 June 1997 (06.06.97)	US	60/057,654	5 September 1997 (05.09.97)	US
60/048,899	6 June 1997 (06.06.97)	US	60/057,646	5 September 1997 (05.09.97)	US
60/048,877	6 June 1997 (06.06.97)	US	60/057,662	5 September 1997 (05.09.97)	US
60/048,963	6 June 1997 (06.06.97)	US	60/057,650	5 September 1997 (05.09.97)	US
			60/057,661	5 September 1997 (05.09.97)	US
			60/057,647	5 September 1997 (05.09.97)	US
			60/070,923	18 December 1997 (18.12.97)	US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

207 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying

secreted proteins of the human that are not yet known.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies,
5 and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

15 In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

20 In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce
25 a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence
30 of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

35 In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking

as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a
35

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5 The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and
10 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability
15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

 The polypeptide of the present invention can be composed of amino acids joined
20 to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,
25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be
30 branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a
35 nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in melanocytes and, to a lesser extent, in testes, ovary, kidney and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, disorders of neural crest derived cells including pigmentation

disorders associated with the nervous system, and other conditions.

Antibodies and probes for differential identification of the tissues or cell types are useful for a number of disorders of the above tissues or cells, particularly of the skin.

reproductive, and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating disorders that arise from alterations in the number or fate of neural crest derived cells including cancers such as melanoma and defects of the developing reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in infant brain and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders of the brain or lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating or diagnosing disorders associated with abnormal proliferation of cells in the Central nervous system and developing lung.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in breast lymph node and to a lesser extent in ovarian cancer and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune responses such as inflammation or immune surveillance for

tumors. This gene may be important for inflammatory responses associated with tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 236 as residues: Lys-45 to Val-50, Lys-69 to Arg-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune responses including those associated with tumor-induced inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in T-cells and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases involving T-cells such as inflammation, autoimmunity, and cancers including T-cell lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of T-cells and other cells of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating T-cell based disorders

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, autoimmunity, infection, or disorders involving activation of monocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 238 as residues: Asp-19 to Arg-31.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing or treating diseases that result in activation of monocytes including infections, inflammatory responses or autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with terminal deoxynucleotidyltransferase which is thought to be important in catalyzing the elongation of oligo- or polydeoxynucleotide chains.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly those of the blood such as leukemia and deficiencies in neutrophils such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to terminal deoxynucleotidyltransferase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and differential diagnosis of acute leukemia's. Alternatively, this gene may function in the proliferation of neutrophils and be useful as a treatment for neutropenia, for example, following neutropenia as a result of chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The contig exhibits a reasonable homology to the human chorionic gonadotropic (HCG) analogue-GT beta-subunit as disclosed in U.S. Patent No. 5,508,261 and PCT Publication No. WO 92/22568. There is a high degree of conservation of the structurally important cysteine residues in these identities.

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 241 as residues: Ser-14 to Pro-22, Leu-43 to Val-53.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 242 as residues: Tyr-22 to His-35.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth

factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune dysfunctions including cancer of the T lymphocytes and autoimmune disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune disorders particularly of T-cell origin and may act as a growth factor for particular subsets of T-cells such as CD4 positive cells which would make this a useful therapeutic for the treatment of HIV and other immune compromising illnesses.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene is routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene is routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor or differentiation factor for particular cell types in the developing fetus and may be useful in replacement or other types of therapy in cases where the gene is expressed aberrantly.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

10 This gene is expressed primarily in T-cells and to a lesser extent in tumor tissue including glioblastoma, meningioma, and Wilm's tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system including autoimmune conditions such as
15 rheumatoid arthritis, inflammatory disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
20 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 245 as residues:
25 Thr-9 to Ser-14.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/ modulation of immune function disorders, including rheumatoid arthritis and inflammatory responses.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in placenta and to a lesser extent in fetal liver and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of hematological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the hematological and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells in the treatment of chemotherapy patients or kidney disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed primarily in stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematopoietic disorders including cancer, neutropenia, anemia, and thrombocytopenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells, in particular following chemotherapy treatment.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with epsilon-COP from *Bos taurus* which is thought to be important as a component of coatomer, a complex of seven proteins that is the major carrier of transport of materials from the

MAPPAPGPASGGSGEVDLEFDVKNAFYIGSYQQCINEAXXVKLSSPERDVERD

VFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMFADYLAHESRRDSIVAELDRE
 MSRSXDVTNTTFLMAASIYLHDQNPDAALRALHQGDSLECTAMTVQILLKLD
 RLDLARKELKRMQDLDEDATLTQLATAWVSLATGGEKLQDAYYIFQEMADKCS
 PTLNLLNGQAACHMAQGRWEAAEGLLQEALDKDSGYPETLVNLIVLSQHLGKP
 5 PEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRLVLQYAPSAEAGPELSGP
 (SEQ ID NO:458); or RDVERDVFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMF
 ADYLAHESRRDSIVAELDREMSRSXDVTNTTFLMAASIYLHDQNPDAALRALH
 QGDSLECTAMTVQILLKLDRLDLARKELKRMQDLDEDATLTQLATAWVSLATG
 GEKLQDAYYIFQEMADKCSPTLNLNNGQAACHMAQGRWEAAEGLLQEALDKD
 10 SGYPETLVNLIVLSQHLGKPPEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRL
 VLQYAPSA (SEQ ID NO:459).

This gene is expressed primarily in activated monocytes and T-cells, and to a lesser extent in multiple other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, immunomodulation, specifically relating to transport problems in these
 cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
 providing immunological probes for differential identification of the tissue(s) or cell
 20 type(s). For a number of disorders of the above tissues or cells, particularly of the
 immune, expression of this gene at significantly higher or lower levels may be routinely
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 taken from an individual having such a disorder, relative to the standard gene
 25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution and homology to epsilon-COP indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for treating
 /diagnosing problems with the cellular transport of proteins that may result in
 30 immunologic dysfunction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with an RNA
 helicase which is thought to be important in polynucleotide metabolism. The translation
 35 product of this contig exhibits good homology to the LbeIF4A antigen of *Leishmania*
braziliensis. The LbeIF4A antigen, or immunogenic portions of it, can be used to
 induce protective immunity against leishmaniasis, specifically *L. donovani*, *L. chagasi*,

L. infantum, *L. major*, *L. braziliensis*, *L. panamensis*, *L. tropica* and *L. guyanensis*. It can also be used diagnostically to detect *Leishmania* infection or to stimulate a cellular and/or humoral immune response or to stimulate the production of interleukin-12.

This gene is expressed primarily in colon cancer and to a lesser extent in
5 pituitary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers particularly of the colon. Similarly,
10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
15 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 249 as residues: Glu-93 to Ala-98, Gln-150 to Leu-
20 156, Leu-220 to Leu-231, Leu-268 to Arg-273, Val-324 to Pro-341, Arg-372 to Asn-380, Ser-405 to Gly-410, Phe-426 to Ala-433, Glu-458 to Asp-470, Arg-506 to Ser-547.

The tissue distribution and homology to RNA helicase indicates that polynucleotides and polypeptides corresponding to this gene are useful for development
25 of diagnostic tests for colon cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this contig has sequence homology to a cytoplasmic protein that binds specifically to JNK designated the JNK interacting protein-1 or JIP-1
30 in mice. JIP-1 caused cytoplasmic retention of JNK and inhibition of JNK-regulated gene expression.

This gene is expressed primarily in brain including pituitary cerebellum frontal cortex, fetal brain and to a lesser extent in the kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of cancers particularly of the colon. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, the translation product of this contig may suppress the effects of the JNK signaling pathway on cellular proliferation, including transformation by the Bcr-Abl oncogene. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 250 as residues: Pro-6 to Ser-26, Ala-30 to Asp-41, Gly-55 to Ser-61, Gly-74 to Thr-80, Tyr-117 to Ala-123, Tyr-167 to Asp-172, Ala-212 to Cys-223, Pro-239 to Tyr-244.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for enhanced survival and/or differentiation of neurons as a treatment for neurodegenerative disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The translation product of this gene shares sequence homology with a liver stage antigen from a protozoan parasite.

This gene is expressed primarily in fetal tissue and to a lesser extent in activated T-cells and other immune cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and diseases of immune function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to a protozoan antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/immune modulation of parasitic infections.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

Preferred polypeptide encoded by this gene comprise the following polypeptide sequences:

MKAIGIEPSLATYHHIIRLFDQPGDPLKRSSFIIYDIMNELMGKRFSPKD
 PDDDKFFQSAMSISSLRDLELAYQVHGLLKTGDNWKFIGPDQHRNFYYSKFF
 10 DLICLMEQIDVTLKWIYEDLIPSAYFPHSQTMIHLLQALDVANRLEVIPKIWER
 (SEQ ID NO:460); and/or KDSKEYGHTFRSDLREEILMLMARDKHPPQLQVAF
 ADCAADIKSAYESQPIRQTAQDWPATSLNCIAILFLRAGRTQEAWKMLGLFRKH
 NKIPRSELLNELMDSAKVSNSPSQAIEVVELASAFSLPICEGLTQVRMSDFAINQ
 EQKEALSNTALTSDSDTSSSDSDSDTSEGK (SEQ ID NO:461). Polynucleotides

15 encoding such polypeptides are also provided.

This gene is expressed primarily in stromal and CD34 depleted bone marrow cells and to a lesser extent in tissues of embryonic origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of hematologic origin including cancers and immune dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
 25 the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 252 as residues: Ser-28 to Gln-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells which may be used for the treatment of hematopoietic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

Preferred polypeptide fragments can be found in an alternative open reading frame. These preferred polypeptides comprise the amino acid sequence:

MSSDNESDIEDLDKLELRLRDKHLKEIQDLQSRQKHEIESLYTKLGKVPPAVI
 5 IPPAAPLSGRRRRPTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQOTL
 HPPGNIPESGQNQLLQPLKPSRSSDNLVSAFTSDGAISVPSLSAPGQGTSTNTV
 GATVNSQAAQAQPPAMTSSRKGTFTDDLHKLVDNWARDAMNLSGRRGSKGH
 MNYEGPGMARKFSAPGQLCISMTSNLGGAPISAASATSLGHFTKSMCPPQQY
 GFPATPFGAQWSGTGGPAPQPLGQFQPVGTASLQNFNISNLQKXISNPPGSNL
 10 RTT (SEQ ID NO:462); IQDLQSRQKHEIESLYTKLGKVPPAVIIPPAAPLSGRRRR
 PTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQOTLHPPGNIPESGQN
 QLLQPLKPSRSSDNLVSAFTSDGAISVPSLSAPGQGTST (SEQ ID NO:463);
 TSDGAISVPSLSAPGQGTSTNTV GATVNSQAAQAQPPAMTSSRKGTFTDDLH
 (SEQ ID NO:464); KGHMNYEGPGMARKFSAPGQLCISMTSNLGGAPISAAS
 15 ATSLGHFTK (SEQ ID NO:465); QPLKPSRSSDNLVSAFTSDGAISVPSLSAPG
 (SEQ ID NO:466). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

This gene is expressed in fetal liver and tissues associated with the CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as
 20 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, liver and CNS diseases. Similarly, polypeptides and antibodies directed
 to these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 25 tissues or cells, particularly of the liver and CNS, expression of this gene at
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 30 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 253 as residues:
 Gln-26 to Lys-34.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosis and treatment for liver diseases such
 35 as hepatocellular carcinomas and diseases of the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

In an alternative reading frame, this gene shows sequence homology to two recently cloned genes, karyopherin beta 3 and Ran_GTP binding protein 5. (See Accession Nos. gil2102696 and gnllPIDle328731.) The Ran_GTP binding protein is related to importin-beta, the key mediator of nuclear localization signal (NLS)-dependent nuclear transport. Based on homology, it is likely that this gene may activity similar to the RAN_GTP binding protein. Preferred polypeptide fragments comprise the amino acid sequence: VRVAAAESMXLLLECA XV RGPEYLTQMWHFMCDALIKA IGTEPDSDVLSEIMHSFAK (SEQ ID NO:467). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in thymus tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in prostate and osteoclastoma tissues. Preferred polypeptide fragments also comprise the amino acid sequence: MEINNQNCFIVIDLVRTVMENGVEGLLIFGAFLPESWLIGVRCSSPEPKALLLIL AHSQKRRLDGWSFIRHLRVHYCVSLTIHFS (SEQ ID NO:468). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and prostate systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
5 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 255 as residues: Met-1 to Ser-11.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for bone and prostate disorders, especially cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

15 This gene shares sequence homology with the FK506-binding protein (FKBP-13) family, a known cytosolic receptor for the immunosuppressants. Recently, another group has cloned a very similar gene, recognizing the homology to FK506-binding protein family, calling their gene FKBP23. (See Accession No. 2827255.)

This gene is expressed primarily in lymphoid tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample, especially for those susceptible to immune suppressant therapies and for diagnosis of diseases and conditions, which include, but are not limited to, immune suppressant disorders. Similarly, polypeptides and antibodies directed to these
25 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 256 as residues: Ala-19 to Val-31, Arg-38 to Gly-49, Ala-61 to Lys-66, Tyr-68 to Pro-78, Gly-116 to Ala-121, Asp-154 to
35 Ser-162, Glu-173 to Gln-186, Phe-194 to Gly-203, Pro-207 to Val-212.

The tissue distribution and homology to FKBP-12 and -13 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune suppressant disorders.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 24**

This gene is expressed primarily in the brain and in the retina. This gene maps to chromosome 8, and therefore can be used in linkage analysis as a marker for chromosome 8.

Therefore, polynucleotides and polypeptides of the invention are useful as
 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and ocular associated disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of
 15 disorders of the above tissues or cells, particularly of the disorders of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
 20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 257 as residues: Cys-34 to Asp-40.

The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders
 25 including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma. Expression in the brain indicates a role in the is useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive
 30 disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene shows sequence homology to a newly identified class of proteins expressed in the nervous system, called *Leukemia Inhibitory Factor* (LIF).

LIF is a glycoprotein involved as a key component in various intracellular signaling pathways. These pathways affect cell proliferation and differentiation.

Preferred polypeptide fragments comprise the amino acid sequence:

QDKHAEVRKNKELKEEASR (SEQ ID NO:469); QQDLSPWAAPVGCPLXXASX
TCHXLPLSGCLRRQSXSLPVVAXLCFWFSCPLASLFVPGQPCVTCPFPSLPFQD
KHAEVRKNKELKEEASR (SEQ ID NO:470). Also preferred are the

5 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
10 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
20 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

The polynucleotide sequence of this gene contains a domain similar to a Flt3 ligand peptide. Preferred polypeptide fragments comprise the amino acid sequence:
PTRCCTTQPCRSSARRPCWVPMVPSPEGREXQPTCPS (SEQ ID NO:471). Thus,
30 this gene may have activity as binding to Flt3 receptors, a process known to promote angiogenesis and/or lymphangiogenesis.

This gene is expressed in human tonsil, and to a lesser extent in teratocarcinoma, placenta, colon carcinoma, and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample
35 and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the tonsil, as well as cancers, such as colon, reproductive, and kidney cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful

in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tonsils, colon, reproductive organs, and kidneys, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 259 as residues: Pro-22 to Glu-33.

The tissue distribution in tonsil and several cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the tonsil or colon, such as tonsillitis, inflammatory diseases involving nose and paranasal sinuses, especially during the infection of influenza, adenoviruses, parainfluenza, rhinoviruses. The gene may also be useful in the diagnosis and treatment of neoplasms of nasopharynx or colon origins.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In an alternative reading frame exists a large open reading frame that encodes a preferred polypeptide. Preferred polypeptide fragments comprise the amino acid sequence:

MKRSLNENSARSTAGCLPVPLFNQKKRNRQPLTSNPLKDDSGISTPSDNYDFP
 PLPTDWAVEAVNPEXAPVMKTVDTGQIPHVSVRPLRSQDSVFENSIQSNTGRSQ
 GGWSYRDGKNKNTSLKWTXKNDFKPQCKRTNLVANDGKNSCPMSSGAQQQK
 QLRTPEPPNLSRNKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNFQQNQY
 KXQMLDDIPEDNTLKETSLYQLQFKEKASSLRIISAVIESMKYWREHAQKTVLL
 FEVLAVLDSAVTPGPYYSKTFLMRDGKNTLPCVFYEIDRELPRILIRGRVHRCVG
 NYDQKKNIFQCVSVRPASVSEQKTFQAFVKIADVEMQYYINVMNET (SEQ ID
 NO:472); SQDSVFENSIQSNTGRSQGGWSYRDGKNKNTSLKWTXKNDFKPQCKR
 (SEQ ID NO:473); NKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNF (SEQ ID
 NO:474); SSLRIISAVIESMKYWREHAQKTVLLFEVLAVLDSAVTPGPYYSKTFLM
 (SEQ ID NO:475); and PRLIRGRVHRCVGNYDQKKNIFQCVSVRPASVSEQKTFQAFV (SEQ ID NO:476).

This gene is expressed in nasopharynx and colon.

use of the differential identification of the tissues or cell types is present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, male reproductive disorders, including cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system,
5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
10 disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a hormone with reproductive or other systemic functions; contraceptive development; male infertility of testicular causes, such as Klinefelter's syndrome, varicocele, orchitis; male sexual dysfunctions; testicular
15 neoplasms; and inflammatory disorders such as epididymitis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as
20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases relating to T cells, as well as cancer in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
25 a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
30 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders. Moreover, since the gene was isolated from an apoptotic cell and based on the understanding of the relationship
35 of apoptosis and cancer, it is likely that this gene may play a role in the genesis of cancer.

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as

5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
10 tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level
15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal diseases.

The translation product of this gene shares sequence homology with C44C1.2 product of *Caenorhabditis elegans* with unknown function. Preferred polypeptide motifs comprise the amino acid sequence:

GVFRPCVCCGRPASLTCSPLDPEVGPYCDTPTMRTLFLNLLWLALACSPVHTTLSK
25 SDAKKAASKTLLEKSQFSDKPVQDRGLVVTDLKAESVVLEHRSYCSAKARDRH
FAGDVLGYVTPWNSHGYDVTKVFGSKFTQISPVWLQLKRRGREMF EVTGLHD
VDQGWMR A VRKHAKGLHIVPRLLFEDWTYDDFRNVLDSEDEIEELSKTVVQVA
KNQHFDG FVVEVWNQLLSQKRVLHIMLTHLAEALHQARLLALLVIPPAITPGT
DQLGMFTHKEFEQLAPVLDGFSLMTYDYSTAHQPGPNAPLSWVRACVQVLDP
30 KXKWRTKSSWGSTSMXWTXRXPDARXPVVGXRXIQXLKDHXPRMVLDISK
PQ (SEQ ID NO:477); TCSPLDPEVGPYCDTPTMRTLFLNLLWLALACSPVHTTLS
(SEQ ID NO:478); LVVTDLKAESVVLEHRSYCSAKARDRH FAGDVLGYVTPW
NSHGYDVTKVFGSKF (SEQ ID NO:479); REMFEVTGLHIDVDQGWMR A VRK
HAKGLHIVPRLLFEDWTYDDFRNVLDSEDE (SEQ ID NO:480); HLTGCLLTLA

Also preferred are polynucleotide fragments encoding these

polypeptide fragments. This gene maps to human chromosome 11, and therefore is useful in linkage analysis as a marker for chromosome 11.

This gene is expressed primarily in human T cells and to a lesser extent in human colon carcinoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 263 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders and gastrointestinal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with Ribosomal protein L11 of *Caenorhabditis elegans*. (See Accession No. 156201.) Preferred polypeptide fragments comprise the amino acid sequence:

ERGVSINQFCKEFNERTKDIKEGIPLPTKILVKPDRTFEIKIGQPTVSYFLKAAAG
IEKGARQTGKEVAGLVTLKHVYEIARIKAQDEAFALQDVPLSSVVRISIIGSARSL
GIRVVKDLSSEELAAFQKERAIFLAAQKEADLAAQEEAAKK (SEQ ID NO:483).

Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in human embryo tissue and to a lesser extent in human epithelioid sarcoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development disorders and epithelial cell cancer. Similarly, polypeptides and antibodies

directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic and epithelial cell systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 264 as residues: Lys-34 to Gly-40.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental disorders and epithelial cancer.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

This gene is expressed primarily in resting T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders of immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is believed to reside on chromosome 11A and is 1.5 kb in length.

This gene is expressed in a variety of tissues including the placenta and foetus, embryonic brain, human umbilical vein endothelial cells, and amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary system and nervous system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of disorders of the urinary and nervous systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene shares sequence homology with R05G6.4 gene product. (See Accession No. gi11326338.) This gene also shares sequence homology with the cyclophilin-like protein CyP-60. (See Accession No. 1199598, see also Biochem. J. 314 (1), 313-319 (1996).) Preferred polypeptide fragments comprise the amino acid sequence:

AVYTYHEKKKDTAASGYGTQNIRLSRDAVKDFDCCCLSLQPCHDPVVTPDGYL
YEREAILEYILHQKKEIARQMKAYEKQQRGTRREEQKELQRAASQDHVRGFLEKE
SAIVSRP LNPFTAKALSGTSPDDVQPGPSVGPPSKDKDKVLPSFWIPSLTPEAK
ATKLEKPSRTVTCMSGKPLRMSDLTPVHFTPLDSSVDRVGLITRSEYVCAVT
RDSLSNATPCAVLRPSGAVVTLECVEKLIRKDMVDPVTGDKLTDRDIIVLQRG
(SEQ ID NO:484); YLYEREAILEYILHQKKEIARQMKAYEKQQRGTRREEQKELQ
RAASQDHVRGFLE (SEQ ID NO:485); and FTAKALSGTSPDDVQPGPSVGPP
SKDKDKVLPSFWIPSLTPEAKATKLEKPSRTVTCMSGKPL (SEQ ID NO:486).

Also preferred are polynucleotide fragments that encode these polypeptide fragments.

This gene is expressed primarily in human testis and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders and in particular testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system. Expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the male reproductive system and in particular of testicular cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

The translation product of this gene shares sequence homology with Lpe5p of *Saccharomyces cerevisiae* which is thought to be important in the metabolism of phospholipids.

This gene is expressed primarily in liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 268 as residues: Pro-14 to Leu-20, Lys-28 to Asn-38, Arg-109 to Arg-114, Lys-119 to Asn-124, Glu-152 to Leu-157, Pro-172 to Val-180.

The tissue distribution and homology to Lpe5p of *Saccharomyces cerevisiae* indicates that polynucleotides and polypeptides of the invention are useful for

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene shares sequence homology with the nuclear ribonucleoprotein U (HNRNP U), encoded by *C. elegans* (See Accession gill703576.) Preferred polypeptide fragments comprise the amino acid sequence:

- 5 MDTSENRPENDVPEPPMPIADQVSNDDRPEGSVEDEEKKESSLPKSFKRKISVV
SATKGVPAAGNSDTEGGQPGRKRRWGASTATTQKKPSISITTESLKSLIPDIKPL
AGQEAVVDLHADDSEDETERNGDDGTHDKGLKICRTVTQVVPVPAEQENGQ
REEEEEKEPEAEPPVPPQVSVEVALPPPAEHEVKKVTLGDTLTRRSISQKSGV
SITIDDPVRTAQVPSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAFWI
10 DKKSHCFVTYSTVEEAVATRTALHGVKWPQSNPKFLCADYAEQDELDYHRGL
LVDRPSETKTEEQGIPRLHPPPPPPVQPPQHPRAEQREQERAVREQWAERERE
MERRERTRSEREWDKVRREGPRSRSRXRRRKERAKSKEKKSEKKEKAQE
EPPAKLLDDLFRKTKAAPCIYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEE
EQKEREKEAERERNRQLEREKRREHSRERDRERERERDRGDRDRDRERDRE
15 RGRERDRRDTKRHSRSTPVRDRGGR (SEQ ID NO:488). Also preferred are
the polynucleotide fragments encoding this polypeptide fragments.

This gene is expressed primarily in epididymus.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, diseases of the male reproductive system. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the male reproductive system, expression of
25 this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the diagnosis and treatment of male
reproductive disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory diseases and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene shares sequence homology with human opsonin protein P35 fragment. (See Accession No. R94181.) The opsonin protein activates the phagocytosis of pathogenic microbes by phagocytic cells. Preferred polypeptide fragments comprise the amino acid sequence: GCDSCPPHLPREAFAQDTQAEGECSSRAERADMCPDAP PSQEVPEGPGAAP (SEQ ID NO:489). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in immune-related tissues such as thymus, macrophage, T cells and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and infectious disease, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 271 as residues: Lys-9 to Arg-14, Met-38 to Asp-51.

- 5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, as well as the treatment and/or diagnosis of infectious disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

- 10 The translation product of this gene shares sequence homology with alpha-2 type I collagen which is thought to be important in tissue repair. (See, e.g., 211607.) Preferred polypeptide fragments comprise the amino acid sequence: PQLPSCGRPWP GTASVFQSHTQGPREDPDPCRAQGSAGTHCPISLSPPRQ (SEQ ID NO:490). Also preferred are the polynucleotide sequences encoding these polypeptide sequences.

- 15 This gene is expressed primarily in the brain and to a lesser extent in the kidney and thymus

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, kidney, and immune disorders. Similarly, polypeptides and
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, kidney, and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution and homology to alpha-2 type I collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tissue repair, and brain, kidney, immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

- 35 The translation product of this gene shares sequence homology with mini-collagen which is thought to be important in tissue repair tumor metastasis. (See Accession No. gnllPID1006976.) Preferred polypeptide fragments comprise the amino acid sequence: PGFRGPSGLGCSFFPRSLGRVLPPGCQRPGAHAD

SSPPPTP (SEQ ID NO:491). Also preferred are polynucleotides encoding this polypeptide fragment.

This gene is expressed in ovarian cancer and to a lesser extent in dendritic cells and smooth muscle.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumor metastasis and tissue repair. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor metastasis and tissue repair, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 273 as residues: Asn-2 to His-11.

20 The tissue distribution and homology to mini-collagen gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumor metastasis and tissue repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene shares sequence homology with the HIV TAT protein. (See
25 Accession No. 328416.) Preferred polypeptide fragments comprise the amino acid sequence: EDLKKPDPASLRAASCGEGKKRKACKNCTCGLAELEEKEK SREQMSSQPKSACGNCYLGD AFRASC PYLGMPAFKPGEKVLLS (SEQ ID NO:492); EDLKKPDPASLRAASCGEGKKRKACKNCTCGLAELEEKEK SREQMSSQPKSACGNCYLGD AFRASC PYLGMPAFKPGEKVLLSDSNLHD
30 (SEQ ID NO:493); CGNCYLGD AFRASC PYLGMPAFKPGEKVLLSDS (SEQ ID NO:494); SCGEGKKRKACKNCTCGLAELEEKE (SEQ ID NO:495); SQPKSAC GNCYLGD AFRASC (SEQ ID NO:496); and REAGQNSERQYVS LSRD (SEQ ID NO:497). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, testes and breast disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, testes and breast disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 274 as residues: Pro-7 to Val-15.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of brain, testes and breast, and other related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in the infant brain, human cerebellum, and to a lesser extent in medulloblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain related disorders and medulloblastoma and other brain cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain related disorders and brain cancers, including medulloblastoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 275 as residues: Thr-41 to Glu-47.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain related disorders, brain cancers, and medulloblastoma.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The translation product of this gene shares sequence homology with a phosphotyrosine-independent ligand for the lck SH2 domain which is thought to be important in signal transduction related to phosphotyrosine-independent ligand for the lck SH2 domain. (See Accession No. gill184951.) Preferred polypeptide fragments
 10 comprise the amino acid sequence: ESSGQARTLADPGPGWPRQQGMCFGSLT
 GLSTTPHGFLTVSAEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAAL
 DTIQYSKH (SEQ ID NO:498). Also preferred are polynucleotide fragments encoding this polypeptide fragment. It is likely that this gene is a new member of a family of
 phosphotyrosine-independent ligands for the lck SH2 domains.

15 This gene is expressed primarily in the placenta and to a lesser extent in endothelial cells and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
 20 not limited to, reproductive, cardiovascular, immune, and infectious diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, reproductive, and immune system, and infectious diseases, expression
 25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
 30 disorder.

The tissue distribution and homology to a phosphotyrosine-independent ligand for the lck SH2 domain indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cardiovascular, reproductive, and immune system diseases, as well as infectious diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

This gene is expressed primarily in the fetal brain, cerebellum and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as
 5 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, neuronal cell related disorders. Similarly, polypeptides and antibodies
 directed to these polypeptides are useful in providing immunological probes for
 differential identification of the tissue(s) or cell type(s). For a number of disorders of
 10 the above tissues or cells, particularly of the neuronal cell related disorders, expression
 of this gene at significantly higher or lower levels may be routinely detected in certain
 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
 individual having such a disorder, relative to the standard gene expression level, i.e.,
 15 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
 NO: 277 as residues: Thr-20 to Gly-28.

The tissue distribution and homology to proline-rich protein genes indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
 20 and treatment of neuronal cell related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with
 precerebellin of human, which is thought to be important in synaptic physiology. (See
 25 Accession No. gil180251.) It has been observed that cerebellin-like immunoreactivity is
 associated with Purkinje cell postsynaptic structures. Thus, it is likely that this gene
 also have synaptic activity. Preferred polypeptide fragments comprise the amino acid
 sequence: QEGSEPVLLEGECLVVCEPGRAAAGGPGGAALGEAPPGRVAFXAV
 RSHHHEPAGETGNGTSGAIYFDQVLVNEGGGFDRASGSFVAPVRGVYSFRFH
 30 VVKVYNRQTVQVSLMLNTWPVISAFAANDPDVTREAAATSSVLLPLDPGDRVSLR
 LRRGXSTGW (SEQ ID NO:499). Also preferred are polynucleotide fragments
 encoding these polypeptide fragments.

This gene is expressed primarily in cerebellum and infant brain. By Northern
 analysis, a single transcript of 2.4 kb was observed in brain tissues.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neuronal cell signal transduction and synaptic physiology. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell signal transduction and synaptic physiology expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene or gene family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed in fetal liver and spleen, and to a lesser extent in bone marrow, umbilical vein, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the immune system, particularly hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 279 as residues: Asp-30 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares sequence homology with a 12 kD nucleic acid binding protein of Feline calcivirus which is thought to be important in viral replication. (See Accession No. 59264)

- 5 This gene is expressed primarily in human cardiomyopathy and to a lesser extent in T helper cells, fetal brain and synovial sarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiomyopathy as well as viral infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 280 as residues: Trp-20 to Cys-26.

- The tissue distribution in cardiomyopathy and homology to viral 12 kD nucleic acid binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cardiomyopathy, including those caused by ischemic, hypertensive, congenital, valvular, or pericardial abnormalities.
- 25 The gene expression pattern may be the consequence or the cause for these conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

- The translation product of this gene shares sequence homology with tumor necrosis factor related gene product which is thought to be important in tumor necrosis, bacterial and viral infection, immune diseases and immunoreactions.

This gene is expressed primarily in colon and to a lesser extent in ovarian and breast cancers.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary or breast origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Tumor necrosis factors indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of cancers of colon, ovary and breast origins, because TNF family members are known to be involved in the tumor development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with mucins, such as epithelial mucin, which is thought to be important in extracellular matrix functions such as protection, lubrication and cell adhesion (See for example Accession No. R68002). Preferred polypeptide fragments comprise the following amino acid sequence: PRSRPALRPGRQRPPSHSATSGVLRPRKKPDP (SEQ ID NO:500). Also preferred are polynucleotide fragments encoding these polypeptide fragments. Moreover, this gene maps to chromosome 22q11.2-qter, and therefore, can be used as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in corpus colosum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors, especially of corpus colosum, as well as metastatic lesions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the corpus colosum and other solid tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having the disorder.

The tissue distribution and homology to mucins indicates that polynucleotides and polypeptides corresponding to this gene are useful for serum tumor markers or immunotherapy targets because tumor cells have greatly elevated level of mucin expression and shed the molecules into the epithelial tissues.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in CD34 depleted buffy coat cord blood and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disorders and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34 depleted buffy coat cord blood and primary dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders. Secreted or cell surface proteins in the above tissue distribution often are involved in cell activation (e.g. cytokines) or molecules involved in cell surface activation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with Interferon induced 1-8 gene encoded polypeptide which is thought to be important in binding to retroviral rev responsive element. Preferred polypeptide fragment comprise the following amino acid sequences: MTLITPSXKLTFXKGKNSWSSRACSSSTLVDP (SEQ ID NO:501). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in CD34 positive cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, retroviral infection, such as AIDS, and other immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 284 as residues: Gln-51 to Trp-62.

The tissue distribution and homology to interferon induced gene 1-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of retroviral infection including HIV. The factor may be involved in viral stability or viral entry into the cells. Alternatively, the virus/factor complex may elicit the cellular immune reaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene shares sequence homology to immunoglobulin lambda chain (See Accession No. 2865484). Therefore it is likely that this gene has activity similar to an immunoglobulin lambda chain. Preferred polypeptide fragments comprise the following amino acid sequence: GHPSPALSIAPSDGSQLPCDEVPIYGEAHVTRYCKKPLTNS HLETEAQSSSL (SEQ ID NO:502). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma and other immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 502 as residues: Gln-51 to Trp-62.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 502 as residues: Gln-51 to Trp-62.

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 285 as residues: Pro-27 to Thr-32.

5 The tissue distribution in Hodgkin's lymphoma and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune
10 functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene has extensive homology to cDNA for Homo sapiens mRNA for the
15 ISLR gene(See Accession No. AB003184). This protein is considered to be a new member of the Ig superfamily and contains a leucine-rich repeat (LRR) with conserved flanking sequences and a C2-type immunoglobulin (Ig)-like domain. These domains are important for protein-protein interaction or cell adhesion, and therefore it is possible that the novel protein ISLR may also interact with other proteins or cells. The ISLR gene
20 was mapped on human chromosome 15q23-q24 by fluorescence in situ hybridization (See Medline Article No. 97468140). Homology to the ISLR gene has been confirmed by another independent group as well (See Accession No. Hs.102171)

This gene is expressed in a number of tissues including human retina, heart, skeletal muscle, prostate, ovary, small intestine, thyroid, adrenal cortex, testis,
25 stomach, spinal cord, fetal lung and fetal kidney tissues, colon, tonsil and stomach cancer, and to a lesser extent in endometrial stromal cells treated with estradiol, breast tissue, synovium, lymphoma, and number of other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary and breast origins. However, due to the wide range of expression in various tissues, protein may play a vital role in the development of cancer in other tissues as well, not just those mentioned above. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, this gene maps to chromosome 15q23-q24, and therefore, can be used as a marker in linkage analysis for chromosome 15.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

Gene has homology to multidrug resistance gene 1 (See Accession No. P06795). Preferred polynucleotide fragments comprise the following sequence:
 GCTTCGTGTCCAACCCCTCTTGCCCTTCGCTGTGTGCCTGGAGCCAGTCCCCA
 CCACGCTCGCGTTTCCTCCTGTAGTGCTCACAGGTCCCAGCACCGATGGCA
 TTCCCTTTGCCCTGAGTCTGCAGCGGGTCCCTTTTGTGCTTCCTTCCCCTCA
 GG TAGCCTCTCTCCCCCTGGGGCCACTCCCGGGGGTGAGGGGGTTACCCCTT
 CCCAGTGTITTTTATTTCCTGTGGGGCTACCCCCAAAGTATTAAAAGTAGCTTT
 GTAA (SEQ ID NO:503). Also preferred are polypeptide fragments encoded by these polynucleotide fragments.

This gene is expressed primarily in lung, esophagus, leukemia (Jurkat cells) and breast cancers and to a lesser extent in macrophages treated with GM-CSF fetal tissues and wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower level than normal is observed.

Polynucleotide fragments of the invention are useful for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower level than normal is observed.

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, due to the high expression level in lung tissue and the proposed function of the multidrug resistance protein 1 gene as the efflux pump responsible for low-drug accumulation in multidrug-resistant cells, protein as well mutants thereof, may also be beneficial as a target for gene therapy, particularly for the chronic patient. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 287 as residues: Met-1 to Lys-16.

The tissue distribution in wide range of cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of cells in active proliferation, such as cancers. The gene products may be used for cancer markers or immunotherapy target.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene maps to the X chromosome.

This gene is expressed primarily in the brain and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders, including sex-linked disorders, of the above tissues or cells, particularly of the neurological, developmental systems, and cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, this gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for this chromosome.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Klinefelter's, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

- 5 The translation product of this gene shares sequence homology with paxillin which is thought to be important in mediating signal transduction from growth factor receptors to the cytoskeleton. Preferred polynucleotide fragments comprise the following sequence: TGGCTCACTGTCTTACAATCACTGCTGTGGAATCATGA
TACCACTTTTAGCTCTTITGCATCTTCCTICAGTGTATTTTTGTTTTTCAAGAGG
10 AAGTAGATTTTAACTGGACAACCTTTGAGTACTGACATCATTGATAAATAAACT
GGCTTGTGGTTTCAA (SEQ ID NO:506). Also preferred are polypeptide fragments encoded by these polynucleotide fragments. More preferably, polypeptide fragments comprise the amino acid sequence: LDELMAHILTEMQAKVAVRAD
AGKKHLPDKQDHIKASLDSMLGGLEQELQDLGIATVPKGHCASCQKPIAGKVI
15 HALGQSWHPEHFVCTHCKEEIGSSPFFERSGLXYCPNDYHQLFSPRCAYCAAP
ILDKVLTAMNQTWHPEHFFCSHCGEVFGEAEGFHEKDKKPYCRKDFLAMFSPK
CGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCELYH
HRRGTLCHGCGQPITGRCISAMGYKFHPEHFVCAFCFLTQLSKGIFREQNDKTY
CQPCFNKLF (SEQ ID NO:507); KASLDSMLGGLEQELQDLGIATVPKGHC
20 ASCQKPIAGKVIHAL (SEQ ID NO:508); CPNDYHQLFSPRCAYCAAPILDKVL
TAMNQTWHPEHFFCSHCGEVFGEAEG (SEQ ID NO:509); DKKPYCRKDFLAM
FSPKCGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCE
L (SEQ ID NO:510); CGQPITGRCISAMGYKFHPEHFVCAFCFLTQLSKGIFRE
QNDKTYCQ (SEQ ID NO:511). Polynucleotide fragments encoding these preferred
25 polypeptide fragments are also contemplated.

This gene is expressed primarily in brain, and to a lesser extent in the developing embryo.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disease states and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above type, the tissue(s) or cell type(s) may be

obtained by various methods, including, but not limited to, analysis of blood, body fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, since this gene shares homology with a gene that maps to chromosome 11, (See Accession No.T87404), gene as well as its translated product may be used for linkage analysis on chromosome 11.

The tissue distribution and homology to paxillin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and or detection of disease states associated with abnormal signal transduction in brain and/or the developing embryo. This would include treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in fetal spleen, brain, and to a lesser extent in six week old embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, neurological disorders, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 290 as residues: Arg-28 to Gly-34.

The expression of this gene in fetal spleen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. In addition the expression of this gene in the early embryo, indicates a key role in embryo development and hence the gene or gene product could be used in the treatment and or detection of embryonic development defects. This would include

treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntintons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with the gene disrupted in the neurodegenerative disease dentatorubal-pallidoluysian atrophy. Moreover a long open reading frame exists in an alternative frame. Preferred polypeptide fragments

10

comprise the following:

MGSSQSVEIPGGGTEGYHVLRVQENSPGHRAGLEPFFDFIVSINGSRLNKDND
 TLKDLLKXNVEKPVKMLIYSSKTLELRETSVTPSNLWGGQGLLGVSIRFCSFD
 GANENVVHVLEVESNSPAALAGLRPHSDYHIGADTMNESEDLSLIETHEAKP
 LKLYVYNTDTDNCREVIITPNSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKIS
 15 LPGQMAGTPITPLKDGFTVEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVSS
 VLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 PHIMPGVGLPELVNPGLPPLPSMPPRNLPGLIAPLPLPSEFLPSFPLVPESSSAASS
 GELLSSLPPTSNA PSDPATTTAKADAASSLTVDVTPPTAKAPTTVEDRVGDSTPV
 SEKPVSAAVDANASESP (SEQ ID NO:512); SVEIPGGGTEGYHVLRVQENSPGH
 20 RAGLEPFFDFIVSINGSRLNKDNDTLKDLLKXNVEKPVKMLIYSSKTLELRETS
 VTPSNLWGGQGLLGVSIRFCSFDGANENVVH (SEQ ID NO:513); ESNPAA
 LAGLRPHSDYHIGADTMNESEDLSLIETHEAKPLKLYVYNTDTDNCREVIITP
 NSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKISLPGQMAGTPITPLKDGFTVEV
 QLSSVNPPSLSPPGTTGIEQSLTG LSIS (SEQ ID NO:514); RIPTRPFEEGKKI
 25 SLPGQMAGTPITPLKDGFTVEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS
 SVLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 APHIMPGVGLPELVNPGLPPLPSMPPRN (SEQ ID NO:516); PGLPPLPSMPPRN
 LPGIAPLPLPSEFLPSFPLVPESSSAASSGELLSSLPPTSNA PSDPATTTAKADAA
 SSLTVDVTPPTAKAPTTVEDRVGDSTPVSEKPVSAAVDAN (SEQ ID NO:517).

30

This gene is expressed primarily in prostate cancer, and to a lesser extent in the pineal glands and in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases associated with the gene.

Antibodies raised against the polypeptides are useful as primary immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the nervous, pulmonary, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 291 as residues: Asn-9 to Leu-14.

The abundance of this gene in the pineal gland and its homology to a gene disrupted in the neurodegenerative disease state Dentatorubral-pallidoluysian atrophy indicates that this gene may be useful in the treatment and/or detection of other neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. The abundance of this gene in fetal lung would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung; that it may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis; and thus the gen or the gene protein encoded by the gene could be used in the detection and/or treatment of these pulmonary disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene primarily in the embryo, indicates the gene plays a key role in embryo development and that the gene or the protein encoded by the gene could be used in the treatment and or detection of developmental defects in the embryo or in infants.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene displays homology to nestin, an intermediate filament protein, the expression of which correlates with the proliferation of Central Nervous System progenitor cells and that is useful in the identification of brain tumors. This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. AA527348).

This gene is expressed primarily in kidney and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the excretory and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 293 as residues: Thr-128 to Asn-135.

The tissue distribution and homology to nestin indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and/or treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, its abundance in kidney indicates that it is useful in the treatment and detection of acute renal failure and other disease states associated with the kidney.

the calcium independent alpha latrotoxin receptor. Preferred polypeptide fragments

comprise the following amino acid sequence:

IYKVFRTAGLKPESCFENIRSCARXXXXXXXXXXXXWIFGVLHVVHASVV
TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPC (SEQ ID NO:518);
WIFGVLHVVHASVVTAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPC

- 5 C (SEQ ID NO:519). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 2213659) The translation product of this gene shares sequence homology with CD 97, a seven transmembrane bound receptor.

This gene is expressed primarily in infant brain and in endothelial cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the neurological and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the
20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 294 as residues: Lys-13 to Leu-21.

The tissue distribution of this gene suggest that it may be useful in the detection and/or treatment of neurodegenerative disease states and behavioral disorders such as
25 Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder, while its expression in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma and immunodeficiency diseases.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in fetal liver and fetal spleen.

- Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 295 as residues: Ser-91 to Lys-98.

The tissue distribution of this gene fetal liver and spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma and immunodeficiency diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

Gene shares homology with human serum amyloid protein. Preferred polypeptide fragments comprise the following amino acid sequence:
ALTRIPPGDWVINVTAVSFAGKTTARFFHSSPPSLGDQARTDPGHIQRRD (SEQ ID NO:520) (See Accession No. W13671). Also preferred are polynucleotide fragments encoding these polypeptide fragments This gene maps to chromosome 9, and therefore, may be used as a marker in linkage analysis for chromosome 9 (See Accession No. AA004342).

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in fetal liver-spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma, and immunodeficiency diseases.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. AA219669).

This gene is expressed specifically in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
15 the above tissues or cells, particularly of the neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the
20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
25 Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Gene shares homology with a yeast protein. Preferred polypeptide fragments
30 comprise the following amino acid sequence: LQEVNITLPENSVWYERYKFDIP VFHL (SEQ ID NO:521). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 1332638)

This gene is expressed primarily in fetal tissue (fetus and fetal liver).

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver disorders and cancers (e.g. hepatoblastoma). Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 298 as residues: Asn-59 to Glu-64.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

Gene has homology with a B-cell surface antigen which may indicate gene plays a role in the immune response, including, but not limited to disorders and infections of the immune system. Preferred polynucleotide fragments comprise the following sequence: TAGCATGTAGCCAGTCGAATAACNTATAAGGACAAAGTGGAGTC CACGCGTGCGGCCGTCTAGACTAGTGGATCCCCCGGCTGCAGGATTTCGGC ACGAG (SEQ ID NO:523). Also preferred are polypeptide fragments encoded by these polynucleotide fragments (See Accession No.T94535). Additionally, this gene shares homology with an interferon-gamma receptor. Preferred polypeptide fragments also comprise the following amino acid sequence: MQGSGSQFRACLLCLCFSCPC SPGGPRWNSRQGGRFPKTCRAISQNLVFKYKTFCPVRYMQPHRSSLCLHFTS YVFILSTWGSRLTYSTDLLKKKKNSRGGPVPIRPKS (SEQ ID NO:522); MQGSGSQFRACLLCLCFSCPCSPGGPRWNSRQGGRFPKTCRAISQNLVFK (SEQ ID NO:524); PVRYMQPHRSSLCLHFTSYVFILSTWGSRLTYSTDLLKKKK NSRGGPVPIRPKS (SEQ ID NO:525); and GEEQRDCSLGWRGVGMRATHCQAA RMFVLESLPKYAGI (SEQ ID NO:526).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders and conditions (immunodeficiencies, cancer, leukemia, hematopoiesis). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 299 as residues: Thr-41 to Gly-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune disorders, immunosuppressive (transplantation) and immunodeficiencies (e.g. AIDS), inflammation and hematopoietic disorders. The expression of this gene in gall bladder would suggest a possible role for this gene product in digestive disorders, particularly of the pancreas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene maps to chromosome 11, and therefore, may be used as a marker in linkage analysis for chromosome 11 (See Accession No. AA011622).

This gene is expressed primarily in a variety of fetal and developmental tissues (e.g. fetal spleen, infant brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, immune or neurological abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 300 as residues: Ser-38 to Ser-43.

- 5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for developmental abnormalities or fetal deficiencies. The detection in infant brain would suggest a role in neurological disorders (both developmental and neurodegenerative conditions of the brain and nervous system, behavioral disorders, depression, schizophrenia, Alzheimer's disease, Parkinson's
- 10 disease, Huntington's disease, mania, dementia). In addition, the detection in spleen would similarly suggest a role in detection and treatment of immunologically mediated disorders (e.g. immunodeficiency, inflammation, cancer, wound healing, tissue repair, hematopoiesis).

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

This gene is expressed primarily in spleen, T-cells, and fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 20 not limited to, immunological deficiencies, including AIDS and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and cardiovascular systems, expression of this gene at significantly higher
- 25 or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, autoimmune disorders, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. The expression of this gene is elevated in certain disorders (e.g., leukemia, lymphoma, autoimmune disorders, immunodeficiencies, immunosuppression, transplantation, hematopoietic disorders, thrombosis).

The expression of this gene is elevated in certain disorders (e.g., leukemia, lymphoma, autoimmune disorders, immunodeficiencies, immunosuppression, transplantation, hematopoietic disorders, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Gene shares homology with a human collagen protein. Preferred polypeptide fragments comprise the following amino acid sequence:

- 5 MPRKTSKCRQLLCSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPGCXSV
SSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHSKSQGE
GQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGGVKVAATTEREPEFKIK
TGKA (SEQ ID NO:527); CSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPG
CXSVPSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHS
10 (SEQ ID NO:528); QGEGQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGG
VKVAATTEREPEFKIKTGKA (SEQ ID NO:529) (See Accession No. 124886). Also
preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cardiovascular disorders. Similarly, polypeptides and antibodies directed
to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
20 tissues or cells, particularly of the cardiovascular system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
25 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 302 as residues:
Pro-32 to Ser-39.

- The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the treatment and diagnosis of cardiovascular
30 disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

The translation product of this gene shares sequence homology with a chicken
single-strand DNA-binding protein. Preferred polypeptide fragments comprise the
35 following amino acid sequence:

MSPRYPGGPRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRM
TPPRGMVPLGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNTNAN

SIPYSSASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPNR
 PNFPMPGSDGPMGGLGGMESHMHMNGSLGSGDMDISISKNSPNNMSLSNQ
 GTPRDDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:530); MSPRYPGG
 PRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRMTPPRGMVP
 5 LGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNPNTNANSIPYSSASP
 GNY (SEQ ID. NO:531); LNALGGPGMPGMNMGPGGGRPWPNPNTNANSIPYSS
 ASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPN (SEQ ID
 NO:532); GPMGGLGGMESHMHMNGSLGSGDMDISISKNSPNNMSLSNQPGTPR
 DDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:533); TCEHSSEAKAFHDY
 10 (SEQ ID NO:534). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments. (See Accession No. 1562534)

This gene is expressed primarily in placenta and to a lesser extent in the fetal heart and a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, developmental abnormalities, fetal deficiencies, and particularly of the
 cardiovascular system. Similarly, polypeptides and antibodies directed to these
 polypeptides are useful in providing immunological probes for differential identification
 20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
 particularly of the reproductive system, expression of this gene at significantly higher or
 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
 another tissue or cell sample taken from an individual having such a disorder, relative to
 25 the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the detection and treatment of developmental
 abnormalities or fetal deficiencies, ovarian and other endometrial cancers, reproductive
 30 dysfunction, cardiovascular disorders, and pre-natal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed primarily in fetal liver and to a lesser extent in the breast and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, liver disorders (including hepatoblastomas) and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The expression in testes and breast indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine and reproductive disorders (e.g. sperm maturation, milk production, testicular and breast cancers).

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. W93595).

This gene is expressed primarily in smooth muscle and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of restenosis, atherosclerosis, stroke, angina, thrombosis, wound healing and other conditions of heart disease. In addition, the expression in brain would suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

Gene shares homology with human stromalin-2. Preferred polypeptide fragments comprise the following amino acid sequence:

QAFVLLSDLLLIFSPQMIVGGRDFLRPLVFFPEATLQSELASF.LMDHVFIQPGDL.

15 GSGA (SEQ ID NO:535); ACSYLLCNPEFTFFSRADFARSQ.LVDLLTDRFQQE

LEELLQVG (SEQ ID NO:536), QKQLSSLRDRMVAFCELCQSC.LSDVDTEIQEQV

ST (SEQ ID NO:537); QVILPALTLVYFSILWTLTHISKSDAS (SEQ ID NO:538);

STHDLTRWELYEPCCQLLQKAVDTGXVPHQV (SEQ ID NO:539). Also preferred

are polynucleotide fragments encoding these polypeptide fragments (See Accession

20 No.R65208) This gene maps to chromosome 7, and therefore, may be used as a marker in linkage analysis for chromosome 7 (See Accession No. D52585).

This gene is expressed primarily in the brain (infant brain, adult brain, pituitary, cerebellum, hippocampus, schizophrenic hypothalamus, amygdala).

Therefore, polynucleotides and polypeptides of the invention are useful as
 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurodegenerative diseases of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
 30 type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having a disorder of the central nervous system.

For example, a polynucleotide or polypeptide of the invention may be used to detect the expression of this gene in a sample taken from an individual having a disorder of the central nervous system.

comprising a sequence shown in SEQ ID NO: 306 as residues: Thr-25 to Lys-36, Lys-55 to Ser-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in the hypothalamus of a human suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the CNS particularly schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, such as schizophrenia expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 307 as residues: Gly-38 to Ala-44.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of schizophrenia and other disorders involving the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

Preferred polypeptides of the invention comprise the following amino acid sequence encoded by this gene:

LAVSTSFICCADISTALPLGSSRPAPAPRHREHEHGHQARPPRLIXTSLMPLSTP
AAQQLLWTQLTPMGGRPGGRHSPPTLHTGPRALPPGPPHPSLHVAALSLLR

(SEQ ID NO:540). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in endometrial tumor and to a lesser extent in amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and immune disorders particularly cancers of those systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 308 as residues: Ser-3 to Arg-9.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune and reproductive disorders particularly cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in kidney cortex and to a lesser extent in early stage human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders such as renal cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 309.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of renal diseases such as cancer of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in kidney medulla.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, metabolic and renal disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
10 differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the metabolic and renal systems, expression of
this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
15 individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study, treatment and diagnosis of metabolic and renal diseases and disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed in chronic synovitis and microvascular endothelium.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, arthritis and atherosclerosis. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the vascular and skeletal systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

35 The tissue distribution indicates that the protein products of this gene are useful
for study, diagnosis and treatment of arthritic and other inflammatory diseases as well
as cardiovascular diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed in resting T-cells and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for the study and treatment of immune diseases such as inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

20 This gene is expressed in a variety of immune system tissues, e.g., neutrophils,
T-cells, and TNF induced epithelial and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, infectious and immune disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the immune and vascular systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising amino acids 1-10, 11-20, 21-30,
31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-110, 111-120, 121-130, 131-140,
141-150, 151-160, 161-170, 171-180, 181-190, 191-200, 201-210, 211-220, 221-230,
231-240, 241-250, 251-260, 261-270, 271-280, 281-290, 291-300, 301-310, 311-320,
321-330, 331-340, 341-350, 351-360, 361-370, 371-380, 381-390, 391-400, 401-410,
411-420, 421-430, 431-440, 441-450, 451-460, 461-470, 471-480, 481-490, 491-500,
501-510, 511-520, 521-530, 531-540, 541-550, 551-560, 561-570, 571-580, 581-590,
591-600, 601-610, 611-620, 621-630, 631-640, 641-650, 651-660, 661-670, 671-680,
681-690, 691-700, 701-710, 711-720, 721-730, 731-740, 741-750, 751-760, 761-770,
771-780, 781-790, 791-800, 801-810, 811-820, 821-830, 831-840, 841-850, 851-860,
861-870, 871-880, 881-890, 891-900, 901-910, 911-920, 921-930, 931-940, 941-950,
951-960, 961-970, 971-980, 981-990, 991-1000, 1001-1010, 1011-1020, 1021-1030,
1031-1040, 1041-1050, 1051-1060, 1061-1070, 1071-1080, 1081-1090, 1091-1100,
1101-1110, 1111-1120, 1121-1130, 1131-1140, 1141-1150, 1151-1160, 1161-1170,
1171-1180, 1181-1190, 1191-1200, 1201-1210, 1211-1220, 1221-1230, 1231-1240,
1241-1250, 1251-1260, 1261-1270, 1271-1280, 1281-1290, 1291-1300, 1301-1310,
1311-1320, 1321-1330, 1331-1340, 1341-1350, 1351-1360, 1361-1370, 1371-1380,
1381-1390, 1391-1400, 1401-1410, 1411-1420, 1421-1430, 1431-1440, 1441-1450,
1451-1460, 1461-1470, 1471-1480, 1481-1490, 1491-1500, 1501-1510, 1511-1520,
1521-1530, 1531-1540, 1541-1550, 1551-1560, 1561-1570, 1571-1580, 1581-1590,
1591-1600, 1601-1610, 1611-1620, 1621-1630, 1631-1640, 1641-1650, 1651-1660,
1661-1670, 1671-1680, 1681-1690, 1691-1700, 1701-1710, 1711-1720, 1721-1730,
1731-1740, 1741-1750, 1751-1760, 1761-1770, 1771-1780, 1781-1790, 1791-1800,
1801-1810, 1811-1820, 1821-1830, 1831-1840, 1841-1850, 1851-1860, 1861-1870,
1871-1880, 1881-1890, 1891-1900, 1901-1910, 1911-1920, 1921-1930, 1931-1940,
1941-1950, 1951-1960, 1961-1970, 1971-1980, 1981-1990, 1991-2000, 2001-2010,
2011-2020, 2021-2030, 2031-2040, 2041-2050, 2051-2060, 2061-2070, 2071-2080,
2081-2090, 2091-2100, 2101-2110, 2111-2120, 2121-2130, 2131-2140, 2141-2150,
2151-2160, 2161-2170, 2171-2180, 2181-2190, 2191-2200, 2201-2210, 2211-2220,
2221-2230, 2231-2240, 2241-2250, 2251-2260, 2261-2270, 2271-2280, 2281-2290,
2291-2300, 2301-2310, 2311-2320, 2321-2330, 2331-2340, 2341-2350, 2351-2360,
2361-2370, 2371-2380, 2381-2390, 2391-2400, 2401-2410, 2411-2420, 2421-2430,
2431-2440, 2441-2450, 2451-2460, 2461-2470, 2471-2480, 2481-2490, 2491-2500,
2501-2510, 2511-2520, 2521-2530, 2531-2540, 2541-2550, 2551-2560, 2561-2570,
2571-2580, 2581-2590, 2591-2600, 2601-2610, 2611-2620, 2621-2630, 2631-2640,
2641-2650, 2651-2660, 2661-2670, 2671-2680, 2681-2690, 2691-2700, 2701-2710,
2711-2720, 2721-2730, 2731-2740, 2741-2750, 2751-2760, 2761-2770, 2771-2780,
2781-2790, 2791-2800, 2801-2810, 2811-2820, 2821-2830, 2831-2840, 2841-2850,
2851-2860, 2861-2870, 2871-2880, 2881-2890, 2891-2900, 2901-2910, 2911-2920,
2921-2930, 2931-2940, 2941-2950, 2951-2960, 2961-2970, 2971-2980, 2981-2990,
2991-3000, 3001-3010, 3011-3020, 3021-3030, 3031-3040, 3041-3050, 3051-3060,
3061-3070, 3071-3080, 3081-3090, 3091-3100, 3101-3110, 3111-3120, 3121-3130,
3131-3140, 3141-3150, 3151-3160, 3161-3170, 3171-3180, 3181-3190, 3191-3200,
3201-3210, 3211-3220, 3221-3230, 3231-3240, 3241-3250, 3251-3260, 3261-3270,
3271-3280, 3281-3290, 3291-3300, 3301-3310, 3311-3320, 3321-3330, 3331-3340,
3341-3350, 3351-3360, 3361-3370, 3371-3380, 3381-3390, 3391-3400, 3401-3410,
3411-3420, 3421-3430, 3431-3440, 3441-3450, 3451-3460, 3461-3470, 3471-3480,
3481-3490, 3491-3500, 3501-3510, 3511-3520, 3521-3530, 3531-3540, 3541-3550,
3551-3560, 3561-3570, 3571-3580, 3581-3590, 3591-3600, 3601-3610, 3611-3620,
3621-3630, 3631-3640, 3641-3650, 3651-3660, 3661-3670, 3671-3680, 3681-3690,
3691-3700, 3701-3710, 3711-3720, 3721-3730, 3731-3740, 3741-3750, 3751-3760,
3761-3770, 3771-3780, 3781-3790, 3791-3800, 3801-3810, 3811-3820, 3821-3830,
3831-3840, 3841-3850, 3851-3860, 3861-3870, 3871-3880, 3881-3890, 3891-3900,
3901-3910, 3911-3920, 3921-3930, 3931-3940, 3941-3950, 3951-3960, 3961-3970,
3971-3980, 3981-3990, 3991-4000, 4001-4010, 4011-4020, 4021-4030, 4031-4040,
4041-4050, 4051-4060, 4061-4070, 4071-4080, 4081-4090, 4091-4100, 4101-4110,
4111-4120, 4121-4130, 4131-4140, 4141-4150, 4151-4160, 4161-4170, 4171-4180,
4181-4190, 4191-4200, 4201-4210, 4211-4220, 4221-4230, 4231-4240, 4241-4250,
4251-4260, 4261-4270, 4271-4280, 4281-4290, 4291-4300, 4301-4310, 4311-4320,
4321-4330, 4331-4340, 4341-4350, 4351-4360, 4361-4370, 4371-4380, 4381-4390,
4391-4400, 4401-4410, 4411-4420, 4421-4430, 4431-4440, 4441-4450, 4451-4460,
4461-4470, 4471-4480, 4481-4490, 4491-4500, 4501-4510, 4511-4520, 4521-4530,
4531-4540, 4541-4550, 4551-4560, 4561-4570, 4571-4580, 4581-4590, 4591-4600,
4601-4610, 4611-4620, 4621-4630, 4631-4640, 4641-4650, 4651-4660, 4661-4670,
4671-4680, 4681-4690, 4691-4700, 4701-4710, 4711-4720, 4721-4730, 4731-4740,
4741-4750, 4751-4760, 4761-4770, 4771-4780, 4781-4790, 4791-4800, 4801-4810,
4811-4820, 4821-4830, 4831-4840, 4841-4850, 4851-4860, 4861-4870, 4871-4880,
4881-4890, 4891-4900, 4901-4910, 4911-4920, 4921-4930, 4931-4940, 4941-4950,
4951-4960, 4961-4970, 4971-4980, 4981-4990, 4991-5000, 5001-5010, 5011-5020,
5021-5030, 5031-5040, 5041-5050, 5051-5060, 5061-5070, 5071-5080, 5081-5090,
5091-5100, 5101-5110, 5111-5120, 5121-5130, 5131-5140, 5141-5150, 5151-5160,
5161-5170, 5171-5180, 5181-5190, 5191-5200, 5201-5210, 5211-5220, 5221-5230,
5231-5240, 5241-5250, 5251-5260, 5261-5270, 5271-5280, 5281-5290, 5291-5300,
5301-5310, 5311-5320, 5321-5330, 5331-5340, 5341-5350, 5351-5360, 5361-5370,
5371-5380, 5381-5390, 5391-5400, 5401-5410, 5411-5420, 5421-5430, 5431-5440,
5441-5450, 5451-5460, 5461-5470, 5471-5480, 5481-5490, 5491-5500, 5501-5510,
5511-5520, 5521-5530, 5531-5540, 5541-5550, 5551-5560, 5561-5570, 5571-5580,
5581-5590, 5591-5600, 5601-5610, 5611-5620, 5621-5630, 5631-5640, 5641-5650,
5651-5660, 5661-5670, 5671-5680, 5681-5690, 5691-5700, 5701-5710, 5711-5720,
5721-5730, 5731-5740, 5741-5750, 5751-5760, 5761-5770, 5771-5780, 5781-5790,
5791-5800, 5801-5810, 5811-5820, 5821-5830, 5831-5840, 5841-5850, 5851-5860,
5861-5870, 5871-5880, 5881-5890, 5891-5900, 5901-5910, 5911-5920, 5921-5930,
5931-5940, 5941-5950, 5951-5960, 5961-5970, 5971-5980, 5981-5990, 5991-6000,
6001-6010, 6011-6020, 6021-6030, 6031-6040, 6041-6050, 6051-6060, 6061-6070,
6071-6080, 6081-6090, 6091-6100, 6101-6110, 6111-6120, 6121-6130, 6131-6140,
6141-6150, 6151-6160, 6161-6170, 6171-6180, 6181-6190, 6191-6200, 6201-6210,
6211-6220, 6221-6230, 6231-6240, 6241-6250, 6251-6260, 6261-6270, 6271-6280,
6281-6290, 6291-6300, 6301-6310, 6311-6320, 6321-6330, 6331-6340, 6341-6350,
6351-6360, 6361-6370, 6371-6380, 6381-6390, 6391-6400, 6401-6410, 6411-6420,
6421-6430, 6431-6440, 6441-6450, 6451-6460, 6461-6470, 6471-6480, 6481-6490,
6491-6500, 6501-6510, 6511-6520, 6521-6530, 6531-6540, 6541-6550, 6551-6560,
6561-6570, 6571-6580, 6581-6590, 6591-6600, 6601-6610, 6611-6620, 6621-6630,
6631-6640, 6641-6650, 6651-6660, 6661-6670, 6671-6680, 6681-6690, 6691-6700,
6701-6710, 6711-6720, 6721-6730, 6731-6740, 6741-6750, 6751-6760, 6761-6770,
6771-6780, 6781-6790, 6791-6800, 6801-6810, 6811-6820, 6821-6830, 6831-6840,
6841-6850, 6851-6860, 6861-6870, 6871-6880, 6881-6890, 6891-6900, 6901-6910,
6911-6920, 6921-6930, 6931-6940, 6941-6950, 6951-6960, 6961-6970, 6971-6980,
6981-6990, 6991-7000, 7001-7010, 7011-7020, 7021-7030, 7031-7040, 7041-7050,
7051-7060, 7061-7070, 7071-7080, 7081-7090, 7091-7100, 7101-7110, 7111-7120,
7121-7130, 7131-7140, 7141-7150, 7151-7160, 7161-7170, 7171-7180, 7181-7190,
7191-7200, 7201-7210, 7211-7220, 7221-7230, 7231-7240, 7241-7250, 7251-7260,
7261-7270, 7271-7280, 7281-7290, 7291-7300, 7301-7310, 7311-7320, 7321-7330,
7331-7340, 7341-7350, 7351-7360, 7361-7370, 7371-7380, 7381-7390, 7391-7400,
7401-7410, 7411-7420, 7421-7430, 7431-7440, 7441-7450, 7451-7460, 7461-7470,
7471-7480, 7481-7490, 7491-7500, 7501-7510, 7511-7520, 7521-7530, 7531-7540,
7541-7550, 7551-7560, 7561-7570, 7571-7580, 7581-7590, 7591-7600, 7601-7610,
7611-7620, 7621-7630, 7631-7640, 7641-7650, 7651-7660, 7661-7670, 7671-7680,
7681-7690, 7691-7700, 7701-7710, 7711-7720, 7721-7730, 7731-7740, 7741-7750,
7751-7760, 7761-7770, 7771-7780, 7781-7790, 7791-7800, 7801-7810, 7811-7820,
7821-7830, 7831-7840, 7841-7850, 7851-7860, 7861-7870, 7871-7880, 7881-7890,
7891-7900, 7901-7910, 7911-7920, 7921-7930, 7931-7940, 7941-7950, 7951-7960,
7961-7970, 7971-7980, 7981-7990, 7991-8000, 8001-8010, 8011-8020, 8021-8030,
8031-8040, 8041-8050, 8051-8060, 8061-8070, 8071-8080, 8081-8090, 8091-8100,
8101-8110, 8111-8120, 8121-8130, 8131-8140, 8141-8150, 8151-8160, 8161-8170,
8171-8180, 8181-8190, 8191-8200, 8201-8210, 8211-8220, 8221-8230, 8231-8240,
8241-8250, 8251-8260, 8261-8270, 8271-8280, 8281-8290, 8291-8300, 8301-8310,
8311-8320, 8321-8330, 8331-8340, 8341-8350, 8351-8360, 8361-8370, 8371-8380,
8381-8390, 8391-8400, 8401-8410, 8411-8420, 8421-8430, 8431-8440, 8441-8450,
8451-8460, 8461-8470, 8471-8480, 8481-8490, 8491-8500, 8501-8510, 8511-8520,
8521-8530, 8531-8540, 8541-8550, 8551-8560, 8561-8570, 8571-8580, 8581-8590,
8591-8600, 8601-8610, 8611-8620, 8621-8630, 8631-8640, 8641-8650, 8651-8660,
8661-8670, 8671-8680, 8681-8690, 8691-8700, 8701-8710, 8711-8720, 8721-8730,
8731-8740, 8741-8750, 8751-8760, 8761-8770, 8771-8780, 8781-8790, 8791-8800,
8801-8810, 8811-8820, 8821-8830, 8831-8840, 8841-8850, 8851-8860, 8861-8870,
8871-8880, 8881-8890, 8891-8900, 8901-8910, 8911-8920, 8921-8930, 8931-8940,
8941-8950, 8951-8960, 8961-8970, 8971-8980, 8981-8990, 8991-9000, 9001-9010,
9011-9020, 9021-9030, 9031-9040, 9041-9050, 9051-9060, 9061-9070, 9071-9080,
9081-9090, 9091-9100, 9101-9110, 9111-9120, 9121-9130, 9131-9140, 9141-9150,
9151-9160, 9161-9170, 9171-9180, 9181-9190, 9191-9200, 9201-9210, 9211-9220,
9221-9230, 9231-9240, 9241-9250, 9251-9260, 9261-9270, 9271-9280, 9281-9290,
9291-9300, 9301-9310, 9311-9320, 9321-9330, 9331-9340, 9341-9350, 9351-9360,
9361-9370, 9371-9380, 9381-9390, 9391-9400, 9401-9410, 9411-9420, 9421-9430,
9431-9440, 9441-9450, 9451-9460, 9461-9470, 9471-9480, 9481-9490, 9491-9500,
9501-9510, 9511-9520, 9521-9530, 9531-9540, 9541-9550, 9551-9560, 9561-9570,
9571-9580, 9581-9590, 9591-9600, 9601-9610, 9611-9620, 9621-9630, 9631-9640,
9641-9650, 9651-9660, 9661-9670, 9671-9680, 9681-9690, 9691-9700, 9701-9710,
9711-9720, 9721-9730, 9731-9740, 9741-9750, 9751-9760, 9761-9770, 9771-9780,
9781-9790, 9791-9800, 9801-9810, 9811-9820, 9821-9830, 9831-9840, 9841-9850,
9851-9860, 9861-9870, 9871-9880, 9881-9890, 9891-9900, 9901-9910, 9911-9920,
9921-9930, 9931-9940, 9941-9950, 9951-9960, 9961-9970, 9971-9980, 9981-9990,
9991-10000.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, infectious diseases, immune and vascular disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and other immune conditions. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

20 This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammatory and other immune conditions. Similarly, polypeptides and
25 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
30 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 315 as residues:
Ala-83 to Thr-91.

35 The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune and inflammatory system,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, disorders of the inflammatory and immune systems. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
a number of disorders of the above tissues or cells, particularly of the inflammatory and
immune systems, expression of this gene at significantly higher or lower levels may be
30 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune system diseases. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system and inflammatory
system, expression of this gene at significantly higher or lower levels may be routinely
detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene
expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of diseases of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, inflammation and immune system disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the inflammatory and immune system,
expression of this gene at significantly higher or lower levels may be routinely detected
30 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 319 as residues: Met-1 to Gly-6, Gly-32 to Pro-43, Leu-55 to Gln-60.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the immune and inflammatory system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

In specific embodiments, polypeptides of the invention comprise the sequence:
 EQVLALLWPRFELILEMNVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSIN
 5 QTIPNERTMQLLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLME
 RAADDSEVESFQQLLNARTQEFIEELLSPFGLVAFVKEAEALIERGQAERLR
 GEEARVTQLIRGFGSSWKSSVESLSQDVMRSFTNFRNGTSIIQG (SEQ ID
 NO:541). ALLKYRFFYQFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRMLK
 VQYEEVAEKDDLMGVEDTAKKGFXXSKPSRSRNTIFTLTGRGVSISPTELEAPILV
 10 PHTAQR (SEQ ID NO: 542); EQRYPFEALFRSQHYXLLDNSCREYLFICEFFVVS
 GPXAHDLFHAVMGRTLSMTLKHLDSYLADCYDAIAVFLCIHIVLRFRNIAAKRD
 VPALDRYW (SEQ ID NO:543), GGLDTRPHYITRRYAEFSSALVSINQ (SEQ ID
 NO:544); SRKEQLVFLINNYDMMLGVL (SEQ ID NO: 545) and/or ALLKYRFFY
 QFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRMLKVQYEEVAEKDDLMG
 15 VEDTAKKGFXXSKPSLRNTIFTLTGRGVSISPTELEAPILVPHTAQRXEQRYPF
 EALFRSQHYXLLDNSCREYLFICEFFVVS GPXAHDLFHAVMGRTLSMTLKHL
 SYLADCYDAIAVFLCIHIVLRFRNIAAKRDVPALDRYWEQVLALLWPRFELILEM
 NVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSINQTIPNERTMQLLGQLQV
 EVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVL MERAADDSEVESFQQLLN
 20 ARTQEFIEELLSPFGLVAFVKEAEALIERGQAERLRGEEARVTQLIRGFGSSW
 KSSVESLSQDVMRSFTNFRNGTS (SEQ ID NO:546). Polynucleotides encoding
 these polypeptides are also encompassed by the invention. The translation product of
 this gene shares sequence homology with suppressor of actin mutation which is thought
 to be important in mutation suppression.

25 This gene is expressed primarily in fetal liver and to a lesser extent in a variety
 of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 30 not limited to, liver and mutations. Similarly, polypeptides and antibodies directed to
 these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 tissues or cells, particularly of the liver or cancer, expression of this gene at
 significantly higher or lower levels may be compared to the expression level

of a standard gene expression level, e.g., the expression level of a standard gene in a tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 320 as residues: Val-53 to Arg-60, Thr-88 to Thr-94, Ala-142 to Ser-150, Gly-188 to Glu-196, Gly-208 to Ser-214, Thr-227 to Gly-232, Lys-279 to Phe-285.

- 5 The tissue distribution and homology to suppressor of actin mutation suggest that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and of liver disorder or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 88

- 10 This gene maps to chromosome 9, and therefore can be used in linkage analysis as a marker for chromosome 9. In specific embodiments, polypeptides of the invention comprise the sequence:

YEGKEFDYVFSIDVNEGGPSYKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVA
KFIIDNTKGQMLGLGNPSFSDPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYV
15 PGSASMGTTMAGVDPFTGNSAYRSAASKTMNIYFPKKEAVTFDQANPTQILGK
LKELNGTAPEEKKLTEDDLILLEKILSLICNSSSEKPTVQQQLQILWKAINCPEDIV
FPALDILRLSIKHPSVNENFCNEKEGAQFSSHLLNPNKGKPANQLLALRTFC
NCFVGQAGQKLMMSQRESLMSHAIELKSGSNKNI (SEQ ID NO: 547);

- HIALATLALNYSVCFHKD (SEQ ID NO: 548); HNIEGKAQCLSLISTILEVVQ
20 DLEATFRLLVALGTLISDDSNVQLAKS (SEQ ID NO: 549); LGVDSQIKKYSS
VSEPAKVSECCRFILNLL (SEQ ID NO: 550); and/or YEGKEFDYVFSIDVNEGGPS
YKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVAKFIIDNTKGQMLGLGNPSFS
DPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYVPGSASMGTTMAGVDPFTGN
SAYRSAASKTMNIYFPKKEAVTFDQANPTQILGKLKELNGTAPEEKKLTEDDLI
25 LLEKILSLICNSSSEKPTVQQQLQILWKAINCPEDIVFPALDILRLSIKHPSVNENFC
NEKEGAQFSSHLLNPNKGKPANQLLALRTFCNCFVGQAGQKLMMSQRESL
MSHAIELKSGSNKNIHIALATLALNYSVCFHKDHNIEGKAQCLSLISTILEVVQD
LEATFRLLVALGTLISDDSNVQLAKSLGVDSQIKKYSSVSEPAKVSECCRFILN
LL (SEQ ID NO: 551). Polynucleotides encoding these polypeptides are also
30 encompassed by the invention. These polypeptides share significant homology with
phospholipase A2 activating protein which is thought to be important in signal
transduction (see, e.g., Wang et al., Gene 161(2):237-241 (1995)).

- This gene is expressed primarily in endothelial cells, to a less extent in placenta,
endometrial stromal cells, osteosarcoma, testis tumor, muscle, and infant brain that are
35 likely to be rich in blood vessels.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in vascular system, aberrant angiogenesis, tumor angiogenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system or tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in endothelial cells and several potential highly vascularized tissues and its homology to phospholipase A2 activating protein suggest that this gene may be involved in transducing signals for endothelial cells in angiogenesis or vasculogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

In specific embodiments, polypeptides of the invention comprise the sequence: YPNQDGDILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTIS
 20 AYKTPRDKVQCILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLLSTVQYISSFYASCLSGEESYWWMQFTA AVE (SEQ ID NO:552); YPNQDGDILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTISAYKTPRDKVQCILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLLSTVQYISSFYA
 25 SCLSGEESYWWMQFTA AVEFIKTI (SEQ ID NO:553); YPNQDGDILRDQVL (SEQ ID NO:554); EAPWPSAQSEI (SEQ ID NO:555); PVLVFLIKANP (SEQ ID NO:560); SGEESYWWMQFTA AVEFIKTI (SEQ ID NO:556); ADDFVPVLVFLIKANPP (SEQ ID NO:557); YKTPRDKVQCIL (SEQ ID NO:558); and/or GADDFVPVLVFLIK (SEQ ID NO:559). The translation product of this gene shares sequence homology with human ras inhibitor and yeast VPS9p which is thought to be
 30 important in golgi vacuole transport.

This gene is expressed primarily in T cells and melanocytes and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ras inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating signal transduction; diagnosis and treatment of disorders involving T cells and melanocytes.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

This gene maps to chromosome 9 and therefore polypeptides of the invention can be used in linkage analysis as a marker for chromosome 9. The translation product of this gene shares sequence homology with neuronal olfactomedin-related ER localized protein which is thought to be important in influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. In specific embodiments, polypeptides of the invention comprise the sequence: SARASTQPPAGQHGPC (SEQ ID NO:561); MPGRWRWQRDMHPARKLLSLL FLILMGTELTQD (SEQ ID NO:562); SAAPDSLLRSSKGSTRGSL (SEQ ID NO:563); AAIVIWRGKSESRIAKTPGI (SEQ ID NO:564); FRGGGTLVLPPTHT PEWLIL (SEQ ID NO:567); PLGITLPLGAPETGGGD (SEQ ID NO:565); and/or CAAETWKGSQRAGQLCALLA (SEQ ID NO:566).

This gene is expressed in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and endocrinological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 323 as residues: Leu-20 to Ala-26, Arg-32 to Arg-39, Thr-104 to Gly-112.

5 The tissue distribution and homology to olfactomedin-related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for maintenance, growth, or differentiation of neuron cells in pineal gland, therefore, may be useful for diagnosis and treatment of neurological disorders in pineal gland.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in prostate and apoptotic T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate disease and T cell dysfunction. Similarly, polypeptides and
15 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detect abnormal activity in prostate and T cells
25 or probably treatment of this abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in prostate and to a lesser extent in smooth muscle cells, fibroblasts, and placenta.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in prostate or vascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s).

For a number of disorders of the prostate or vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating function of prostate or highly vascularized tissues, e.g. placenta.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 93**

This gene is expressed primarily in embryos and fetal tissues stage human and to a lesser extent in a wide variety of other proliferative tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in embryonic development and cell proliferation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic tissues and proliferative cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of abnormalities in developing and proliferative cells and organs.

30 **FEATURES OF PROTEIN ENCODED BY GENE NO: 94**

The translation product of this gene shares sequence homology with transformation related protein which is thought to be important in transformation.

This gene is expressed primarily in female reproductive tissues, i.e., breast cancer cells, placenta, and ovary and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancer or dysfunction of reproductive tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction system,

- 5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 327 as residues: Ser-50 to Pro-61.

- 10 The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of conditions caused by transformation, i.e. tumorigenesis in reproductive organs, e.g. breast, placenta, and ovary.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in testes, rhabdomyosarcoma, infant brain and to a lesser extent in some tumors and highly vascularized tissues.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumorigenesis, abnormal angiogenesis, and/or neurological disorders. . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor tissues or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 328 as residues: Arg-46 to Trp-54, Pro-60 to Ile-69, Asn-116 to Ala-122, Arg-147 to Leu-153, Ser-158 to Gly-170, and

30 The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for a range of disease states including treatment of

tumor or vascular disorders and the treatment of neurological disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 96**

This gene maps to chromosome 7 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 7. The translation product of this gene is homologous to the *Clostridium perfringens* enterotoxin (CPE) receptor gene product and shares sequence homology with a human
10 ORF specific to prostate and a glycoprotein specific to oligodendrocytes both of which are tissue specific proteins. (See e.g., Katahira et al., J Cell Biol. 136(6):1239-1247 (1997). PMID: 9087440; UI: 97242441.

This gene is expressed primarily in pancreas tumor and ulcerative colitis and to a lesser extent in several tumors and normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic disorder, ulcerative colitis, tumors and food poisoning. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
20 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or tumorigenic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
25 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 329 as residues: Gly-147 to Met-152, Cys-177 to Lys-188.

30 The tissue distribution and homology to prostate and oligodendrocyte-specific protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis or treatment of disorder in pancreas, ulcerative colitis, and tumors. Furthermore, identity to the human receptor for *Clostridium perfringens* enterotoxin indicates that the soluble portion of this receptor could be used in the
35 treatment of food poisoning associated with *Clostridia perfringens* by blocking the activity of *perfringens* enterotoxin.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

The translation product of this gene shares sequence homology with ATPase which is thought to be important in metabolism.

- 5 This gene is expressed primarily in testes and several hematopoietic cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 330 as residues: Leu-37 to Ala-42.

The tissue distribution and homology to ATPase indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis and treatment of leukemia and other hematopoietic disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

In specific embodiments, polypeptides of the invention comprise the sequence: MRSARPSLGCLPSWAFSQALNI (SEQ ID NO:568); LLGLKGLAPAEISAVCE KGNFN (SEQ ID NO:569); VAHGLAWSYYIGYLRILPELQARIR (SEQ ID NO:570); TYNQHYNNLLRGAVSQRC (SEQ ID NO:571); ILLPLDCGVPDNLSM ADPNIRFLDKLPQQTGDRAGIKDRVYSN (SEQ ID NO:572); SIYELLENGQRAGT CVLEYATPLQTLFAMSQYSQAGFSGEDRLEQ (SEQ ID NO:573); AKLFCRTLE DILADAPESQNNCRLIAYQEPADDSSFSLSQEVLRHLRQEEKEEVTVGSLKTSAPV PSTSTMSQEPPELLISGMEKPLPLRTDFS (SEQ ID NO:574); and/or LLGLKGLA PAEISAVCEKGNFNVAHGLAWSYYIGYLRILPEL (SEQ ID NO:575).

This gene is expressed primarily in prostate BPH and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, benign prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urinary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 331 as residues: Ile-60 to Asn-69, Leu-106 to Asp-112, Glu-130 to Gly-136, Phe-160 to Glu-167, Pro-184 to Cys-190, Glu-197 to Ser-202, Arg-215 to Glu-221, Thr-237 to Pro-242.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of benign prostatic hypertrophy or prostate cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders or injuries of the salivary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders of, or injuries to the salivary gland or other glandular tissue.

FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene maps to chromosome 15, accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 15. The translation product of this gene shares sequence homology with a *C.elegans* gene of unknown function. In specific embodiments, polypeptides of the invention comprise the sequence: DPRVRLNSLTCKHIFISLTQ (SEQ ID NO:583); TMKLLKLRRNIV KLSLYRHFTN (SEQ ID NO:576); TLILAVAASIVFIWTTMKFRI (SEQ ID NO:577); VTCQSDWRELWVDDAIWRLLFSMILFVI (SEQ ID NO:578); MVLWR PSANNQRFAFSPLSEEEFEDEQ (SEQ ID NO:580); KEPMLKESFEGMKMRS TKQEPNGNSKVNKAQEDDL (SEQ ID NO:584); and/or KWVEENVPSVTDVALP ALLDSDEERMITHFERSKME (SEQ ID NO:582). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thyroid and to a lesser extent in osteoclastoma, kidney medulla, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thyroid dysfunction or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 333 as residues: Lys-107 to Leu-124, Glu-150 to Thr-159, Pro-173 to Asp-179, Ser-192 to Ser-201.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of thyroid dysfunction or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 101

In specific embodiments, polypeptides of the invention comprise the sequence:

IRHELTVLRDTRPACA (SEQ ID NO:585); and/or MDFXMALIYD (SEQ ID NO:586). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney cortex and to a lesser extent in adult
5 brain, corpus colosum, hippocampus, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to
10 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neurological
20 disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 102

In specific embodiments, polypeptides of the invention comprise the sequence:
MQEMMRNQDRALSNLESIPGGYNA (SEQ ID NO:587); LRRMYTDIQEPMLSA
25 AQEQF GGNPF (SEQ ID NO:588); ASLVSNTSSGEGSQPSRTENRDPLPNPWAP
QT (SEQ ID NO:589); SQSSSASSGTASTVGGTTGSTASGTSGQSTTAPNLVPGV
GASMFNTPG MQSLLQQITENPQLMQNMLSAPY (SEQ ID NO:590);
MRSMMQSLSQNPDLAAQMMLNNPLFAGNPQLQE QMRQQLPTFLQQ (SEQ ID
NO:591); MQNPDTLSAMSNPRAMQALLQIQGLQTLATEAPGLIPGFTPGLG
30 ALGSTGGSSGTNGSNATPSENTSPTAGT (SEQ ID NO:592); TEPGHQQFI
QQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALIATGGDINAA
IERLLGSQPS (SEQ ID NO:593); RNPAMMQEMMRNQDRALSNLESIPGGY
NALRRMYTDIQEPMLSAA (SEQ ID NO:594); GNPFASLVSNTSS (SEQ ID
NO:595); ENRDPLPNPWA (SEQ ID NO:595); GKILKDQDTLSQHGIHD (SEQ ID
35 NO:597); GLTVHLVIKTQNRP (SEQ ID NO:598); SELQSQMQRQLLSNPEDM
(SEQ ID NO:599); PEISHMLNNPDIMR (SEQ ID NO:600); and/or
RQLIMANPQMQQLIQRNP (SEQ ID NO:601). Polynucleotides encoding these

polypeptides are also encompassed by the invention.

This gene is expressed primarily in breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 103

The translation product of this gene shares sequence homology with secreted serine proteases and lysozyme C precursor, which is thought to be important in bacteriolytic function. In specific embodiments, polypeptides of the invention comprise the sequence: NLCHVDCQDLLNPILLAGIHC AKRIVS (SEQ ID NO:602); LDGFEGYSLSDWLCLAFVESKFN (SEQ ID NO:603); NENADGSFDYGLFQINSHYWCN (SEQ ID NO:604); and/or NLCHVDCQDLLNPILLAGIHC AKRIVS (SEQ ID NO:605). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Ile-62 to Phe-70, Asn-78 to Asn-84.

The tissue distribution and homology to lysozyme C precursor indicates that polynucleotides and polypeptides corresponding to this gene are useful for boosting the monocyte-macrophage system and enhance the activity of immunoagents.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 104**

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 105

The translation product of this gene shares sequence homology with ARI protein of Drosophila (accession 2058299; EMBL: locus DMARIADNE, accession X98309), which is thought to be important in axonal path-finding in the central nervous system. In specific embodiments, polypeptides of the invention comprise the sequence IREVNEVIQNPAT (SEQ ID NO:606); ITRILLSHFNWDKEKLMERYF DGNLEKLFA (SEQ ID NO:607); NTRSSAQDMPCQICYLNYPNSYF (SEQ ID NO:608); TGLECGHKFCMQCWSEYLTTKIMEEGMGQTISCPAHG (SEQ ID NO:614); CDILVDDNTVMRLITDSKVKLKYQHLITNSFVECNRLKWCAPD CHHVVKVQYPDAKPV (SEQ ID NO:609); CDILVDDNTVMRLITDSK

VKLKYQHLITNSFVECNRLCLKWCPAPDCHHVVKV (SEQ ID NO:610);
 GCNHMVCRNQNCKAFCWVCLGPWEPHGSAWYNCNRYNEDDAKAARDAQE
 RSRAALQRYL (SEQ ID NO:611); FYCNRYMNHMQSLRFEHKLYAQVKQ
 KMEEMQQHNMSWIEVQFLKKAVDVLCQCRATLMYT (SEQ ID NO: 612);
 5 YVFAFYLLKKNNQSIIFENNQADLENATEVLSGYLERDISQDSLQDIKQKVQDKY
 RYCESR (SEQ ID NO:613) Polynucleotides encoding these polypeptides are also
 encompassed by the invention.

This gene is expressed primarily in adult brain, and to a lesser extent in
 endometrial tumor, melanocytes, and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases or injuries involving axonal path development. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 15 immunological probes for differential identification of the tissue(s) or cell type(s). For
 a number of disorders of the above tissues or cells, particularly of the central nervous
 system, expression of this gene at significantly higher or lower levels may be routinely
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 20 taken from an individual having such a disorder, relative to the standard gene
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution and homology to ARI protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for treatment of
 25 disease states or injuries involving axonal path development, including
 neurodegenerative diseases and nerve injury.

FEATURES OF PROTEIN ENCODED BY GENE NO: 106

The translation product of this gene shares sequence homology with cytochrome
 30 b561 [Sus scrofa] which is thought to be an integral membrane protein of
 neuroendocrine storage vesicles of neurotransmitters and peptide hormones.

This gene is expressed primarily in frontal cortex and to a lesser extent in
 rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of this gene are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 339 as residues: Ser-18 to Pro-24.

The tissue distribution and homology to cytochrome b561 [*Sus scrofa*] indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of neurological disorders. This gene may also be important in regulation of some types of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 107

In specific embodiments, polypeptides of the invention comprise the sequence: MWGYLFVDAAWNFLGCLICGW (SEQ ID NO:615); MHFISSGNVSAIRSSILLRXSLSYLGNCRLRVSAIFVYFLLFLLLS (SEQ ID NO:616); and/or MDQALRGSPSEGFSTDPSPPQVGRQIPSFPPWRRLVLPKASGCFLEREWLVCVFKLRTRPGAEA HAYNSSILGGRGKGIT (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pancreas tumor and to a lesser extent in cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 340 as residues: Pro-22 to Phe-33.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pancreatic tumors.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene maps to chromosome 17 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

10 MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRKRMEKEVSDFIQDSGQIK
KKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYVMIFKKEFAPSDEELD
SYRRGEEWDPQKAEEKRNXXKELAQRRQ (SEQ ID NO:618); EEEAAQQGPVVV
SPADYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIIEAMNE
IRAKKRLRQSGE (SEQ ID NO:619); PPRPAQLPLTPGAGQGAGRDKAAAIRA
15 HPGAPPLNHLLP (SEQ IDNO:620); AVPQAGGKQVFDLSPELGYVRGMCVCV
(SEQ ID NO:621) and/or MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRK
RMEKEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYV
MIFKKEFAPSDEELDSYRRGEEWDPQKAEEKRNXXKELAQRRQEEAAQQGPVVV
SPADYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIIEAMNE
20 IRAKKRLRQSGE (SEQ ID NO:622). Polynucleotides encoding these polypeptides
are also encompassed by the invention. The translation product of this gene shares
sequence homology with FSA-1 which may play a role as a structural protein
component of the acrosome.

This gene is expressed primarily in fetal kidney and sperm.

25

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders, especially involving acrosomal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
30 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or saliva).

individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 341 as residues: Glu-8 to Asn-35.

The tissue distribution and homology to FSA-1 indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of infertility due to
5 acrosomal disfunction of sperm.

FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in pituitary and to a lesser extent in epididymus.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for
15 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
20 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 342 as residues: Met-1 to Trp-6.

Because the gene is found in both pituitary and epididymus, this indicates that
25 polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of male reproductive disorders. This may involve a secreted peptide produced in the pituitary targeting the epididymus.

FEATURES OF PROTEIN ENCODED BY GENE NO: 110

30 In specific embodiments, polypeptides of the invention comprise the sequence: LLCPVLNSGXSWNFPHPSPQEYSFHGFHSTRLWI (SEQ ID NO:623); and/or PSTPWFLFLLGLTCPFSTSHPRWDSIPP (SEQ ID NO:624). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in resting T-cells. .

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, T-cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of certain immune disorders, especially those involving T-cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 111

This gene is expressed primarily in cerebellum and whole brain and to a lesser extent in infant brain and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 344 as residues: Asp-48 to Gly-55.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

This gene encodes a protein which shares sequence homology with a mitochondrial ribosomal protein homologous to ribosomal protein s15 of E. coli which

is thought to be important in the early assembly of ribosomes (See Accession No. M38016). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in developmental tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development of cancers and tumors in addition to healing wounds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribosomal protein s15 of *E. coli* indicates that polynucleotides and polypeptides corresponding to this gene are useful for diseases related to the assembly of ribosomes in the mitochondria which is important in the translation of RNA into protein. Therefore, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of multiple tumors as well as in healing wounds which are thought to be under similar regulation as developmental tissues. Protein, as well as, antibodies directed against the protein have utility as tumor markers, in addition to immunotherapy targets, for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 113

The translation product of this gene shares sequence homology with human poliovirus receptor precursors which are thought to be important in viral binding and uptake. Preferred polypeptide fragments comprise the following amino acid sequence:
 ELSISISNVALADEGEYTCSIFTMPVRTAKSLVTVLGIPQKPIITGYKSSLREKDT
 ATLNCQSSGSKPAARLTWRKGDQELHGEPTRIQEDPNGKTFTVSSSVTFQVTR
 EDDGASIVCSVNHESLKGADRSTSQRIEVLYTPTAMIRPDPPHPREGQKLLHHC
 EGRGNPVPQQYLWEKEGSVPPLKMTQESALIFPFLNKSDSGTYGCTATSNMGS
 YKAYYTLNVND (SEQ ID NO:625). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gnllPID1002627).

This gene is expressed almost exclusively in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, susceptibility to viral disease and diseases of the CNS especially cancers of that system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 346 as residues: Leu-26 to Asp-37, Lys-53 to Ser-59.

The tissue distribution and homology to poliovirus receptor precursors indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and prevention of diseases that involve the binding and uptake of virus particles for infection. It might also be helpful in genetic therapy where the goal is to insert foreign DNA into infected cells. With the help of this protein, the binding and uptake of this foreign DNA might be aided. In addition, it is expected that over expression of this gene will indicate abnormalities involving the CNS, particularly cancers of that system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene shares sequence homology with YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* in addition to alpha-1 collagen type III (See Accession No. gil537432). One embodiment for this gene is the polypeptide fragment(s) comprising the following amino acid sequence: VPELPDRVHQLHQA VQGCALGRPGFPGGPTH SGHHKSHPGPAGGDYNRC DRPGQVHLHNPRGTGRRGQLHPTAGPGVHRRACPSQQLPHRLGPGVPCPSPSLTPVLPSWTQSWCG LPGYTSSS (SEQ ID NO:630). An additional embodiment is the polynucleotide (Gene No. 114).

The gene is expressed in B and T cells and in various other types of cells, including B and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegeneration and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 347 as residues: Glu-34 to Glu-39, Gly-51 to Ser-72, Ala-88 to Glu-93, Gln-100 to Val-105.

The tissue distribution and homology to YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* as well as to a conserved alpha-1 collagen type III protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons' Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 115

The translation product of this gene shares sequence homology with alpha 3 type IX collagen which is thought to be important in hyaline cartilage formation via its ability to uptake inorganic sulfate by cells (See Accession No. gil975657). One embodiment of this gene is the polypeptide fragment comprising the following amino acid sequence: SLRRPRSAAXQTLTTLSSVSSASSSALPGSREPCDPRAPPPR SGSAASCCSCCSCPRRRAPLRSRPGSKRRIRQREVVDLYNGMCLQGPGVPG RDGSPGANGIPGTPGIPGRDGFKEGKEGLRESFEESWTPNYKQCSWSSLNY GIDLGKIAECTFTKMRSNSALRVLFSGSLRLKCRNACCQRWYFTFNGAECSPG LPIEAIYLDQGSPEMNSTINIHRSSVEGLCEGIGAGLVDVAIWVGTCSDYPKG DASTGWNSVSRIIIIEELPK (SEQ ID NO:634). An additional embodiment are the

polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in smooth muscle and to a lesser extent in synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e., spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 3 type IX collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases associated with the mutation in this gene which leads to the many different types of chondrodysplasias. By the use of this product, the abnormal growth and development of bones of the limbs and spine could be routinely detected or treated in utero since the protein or muteins thereof could affect epithelial cells early in development and later the chondrocytes of the developing craniofacial structure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 116

The translation product of this gene shares sequence homology with retrovirus-related reverse transcriptase which is thought to be important in viral replication. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: TKKENC RPASLMNIDTKILNKILMNQ (SEQ ID NO:640). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. pirlA25313IGNHUL1).

This gene is expressed primarily in human meningioma

biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, retroviral diseases such as AIDS, and possibly certain cancers due to transactivation of latent cell division genes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to retrovirus-related reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of diseases and maladies associated with retroviral infection since a functional reverse transcriptase (RT) or RT-like molecule is an integral component of the retroviral life cycle.

FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with an unknown gene from *C. elegans*, as well as weak homolog with mammalian metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is known to be required for embryonic development. Preferred polypeptide fragments comprise the following amino acid sequence: MCNLPKVVCRANAEYMSPSGKVPXXHVGNNQ VVSELGPIVQFVKAKGHSLSDDGLEEVQKAEMKAYMELVNNMLLTAEVLYLQWC DEATVGXITHXRYGSPYPWPLXHILAYQKQWEVKRKXKAIGWGKKTLDQVLE DVDQCCQALSQRLGTQPYFFNKQPTELDALVFGHLYTILTTQLTNDELSEKVKN YSNLLAFCRRI EQHYFEDRGKGRLS (SEQ ID NO:641); MCNLPKVVCRANAE YMSPSGKVPXXHVGNNQVVSELGPIVQFVK (SEQ ID NO:642). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gil1326108).

This gene is expressed primarily in fetal tissues and to a lesser extent in hematopoietic cells and tissues, including spleen, monocytes, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer; lymphoproliferative disorders; inflammation; chondrosarcoma, and Gaucher disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and embryonic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 118

The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA chain from an RNA molecule, and is a method whereby the infecting RNA chains of retroviruses are transcribed into their DNA complements. One embodiment for this gene is the polypeptide fragment comprising the following amino acid sequence:
 MXXXNSHITIFTLNVNGLNAPNERHRLANWIQSQDQVCCIQETHLTGRDTHRL
 KIKGWRKIYQANGKQKK (SEQ ID NO:647). An additional embodiment is the polynucleotide fragments comprising polynucleotides encoding these polypeptide fragments (See Accession No. gil2072964).

This gene is expressed primarily in skin and to a lesser extent in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, hematopoietic disorders; inflammation; disorders of immune surveillance. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential diagnosis of cancer.

For a number of disorders of the above tissues or cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and

wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for cancer therapy. Expression in the skin also indicates that this gene is useful in wound healing and fibrosis. Expression by neutrophils also indicates that this gene product plays a role in inflammation and the control of immune surveillance (i.e. recognition of viral
10 pathogens). Reverse transcriptase family members are also useful in the detection and treatment of AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 119

15 The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA copy of an RNA molecule, and is a method whereby a retrovirus reverse-transcribes its genome into an inheritable DNA copy.

 This gene is expressed primarily in the frontal cortex of brain.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
25 of the above tissues or cells, particularly of the CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution and homology to reverse transcriptase suggest that this is useful in the treatment of cancer and AIDS. The expression in brain indicates that it plays a role in neurodegenerative disorders and in neural degeneration.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 120

One embodiment of this gene has homology to a hypothetical protein in *Schizosaccharomyces pombe* (See Accession No. 2281980). Another embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:
 5 IYHLHSWIFFHFKRAFCMCFITMKVIAHCSKLRKCXNAQISVFCTTLTASYPT
 (SEQ ID NO:651). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

10 This gene is expressed primarily in adult hypothalamus and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disorders; endocrine function; and vertigo. Similarly,
 15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
 20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 25 corresponding to this gene are useful for the treatment and diagnosis of neurodegenerative disorders; diagnosis of tumors of a brain or neuronal origin; treatments involving hormonal control of the entire body and of homeostasis, behavioral disorders, such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and
 30 panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 121

This gene contains an important binding site for a promoter element and thus regulating its transcription (See Accession No. gi33969). This gene maps to

chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in brain and breast and to a lesser extent in a variety of hematopoietic tissues and cells.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer of the brain and breast; lymphoproliferative disorders; neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these
10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, breast, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancer of the
20 brain, breast, and hematopoietic system. In addition, it may be useful for the treatment of neurodegenerative disorders, as well as disorders of the hematopoietic system, including defects in immune competency and inflammation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with an ATP synthase, a key component of the proton channel that is thought to be important in the translocation of protons across the membrane.

30

This gene is expressed primarily in T cell lymphoma.

35

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, T cell lymphoma. Similarly, polypeptides and antibodies directed to these
polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATP synthase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of defects in proton transport, homeostasis, and metabolism, as well as the diagnosis and treatment of lymphoma. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene maps to chromosome 15, and therefore, may be used as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in a variety of fetal tissues, including fetal liver, lung, and spleen, and to a lesser extent in a variety of blood cells, including eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer (abnormal cell proliferation); T cell lymphomas; and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetus and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of T cell lymphomas.

The gene product is a protein that is involved in cell growth, proliferation, apoptosis, or cell survival. Thus it may be useful in the management and

treatment of a variety of cancers and malignancies. In addition, its expression in blood cells suggest that it may play additional roles in hematopoietic disorders and conditions, and could be useful in treating diseases involving autoimmunity, immune modulation, immune surveillance, and inflammation..

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in placenta and to a lesser extent in pineal gland and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, endocrine, and female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 357 as residues: Leu-69 to Val-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders in development. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in benign prostatic hyperplasia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of benign prostatic hyperplasia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive

system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene
5 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of benign prostatic hyperplasia. Protein, as well as, antibodies directed against the protein may show utility
10 as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in apoptotic T-cells and to a lesser extent in suppressor T cells and ulcerative colitis.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving premature apoptosis, and immunological and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these
20 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
25 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders involving
30 inappropriate levels of apoptosis, especially in immune cell lineages. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia

FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in Raji cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and T cell autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 360 as residues: Asp-23 to Gly-29.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammation and T cell autoimmune disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 128

The translation product of this gene shares sequence homology with an *C. elegans* coding region C47D12.2 of unknown function (See Accession No. gnllPIDe348986). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: EDDGFNRSIHEVILKNITWY SERVLTEISLGSLLILVVIRTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQY AAQRIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDYAQDLNVIEEVIRMM LEIINSCLTNSLHHNPNLVALLYKRDLFEQFRTHPSFQDIMQNIDLVISFFSSRLL QAGS (SEQ ID NO:657); EDDGFNRSIHEVILKNITWYSERVLTEISLGSLLILVV (SEQ ID NO:658); RTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQYAAQ RIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDY AQD (SEQ ID NO:661); SCLTNSLHHNPNLVALLYKRDLFEQFRTHPSFQD IMQNIDLVISFFSSRLLQAGS (SEQ ID NO:660). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to

chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in smooth muscle and to a lesser extent in fetal liver.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, atherosclerosis and other cardiovascular and hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
- 15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of circulatory system
- 20 disorders such as atherosclerosis, hypertension, and thrombosis. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the
- 25 expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 129

- 30 The translation product of this gene shares sequence homology with a ribosomal protein which is thought to be important in cellular metabolism, in addition to the *C.elegans* protein F40F11.1 which does not have a known function at the current time (See Accession No. gnllPIDle244552). Preferred polypeptide fragments comprise the following amino acid sequence:

...GK...LLEKRVLT...SASTKQKHS...VLTQDEDAEDHCHPPRESALHPQV...
PLREAPQEHVCTPVPLLQGRPDR (SEQ ID NO:662); MKMQRTIVIRRDYLIH

YIRKYNRFEKRRHKNMSVHLSPCFRDVQIGDIVTVGECRPLSKTVRFNVLKVTK
 AAGTKKQFQKF (SEQ ID NO:663); MADIQTERAYQKQPTIFQNKKRVLGET
 GK (SEQ ID NO:664); HCHPPRLSALHPQVQPLREAPQEHVCTPVPL LQGRPDR
 (SEQ ID NO:666); NIGLGFKDTPRLLRGTYIDKKCPFTGNVSIRGRILSGVVTO
 5 (SEQ ID NO:669); MKMQRTIVIRRDYLHYIRKYNRFEKRRHKNMSVHLSP (SEQ
 ID NO:667); CFRDVQIGDIVTVGECRPLSKTVRFNVLKVTKAAGTKKQFQKF
 (SEQ ID NO:668). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

10 This gene is expressed primarily in Wilm's tumor and to a lesser extent in
 thymus and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases affecting RNA translation. Similarly, polypeptides and
 15 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the Wilm's tumors, expression of this gene
 at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 362 as residues:
 Thr-11 to Asp-20.

25 The tissue distribution and homology to a ribosomal protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for diseases
 affecting RNA translation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 130

30 The translation product of this gene shares sequence homology with a yeast
 DNA helicase which is thought to be important in global transcriptional regulation (See
 Accession No. gnllPIDe243594). One embodiment for this gene is the polypeptide
 fragments comprising the following amino acid sequence: IFYDSDWNPTVDQQA
 MDRAHRLGQTKQVTYRLICKGTIEERILQRAKEKSEIQRMVISG (SEQ ID
 35 NO:670); TRMIDLLEEYMVYRKHTYXRLDGSSKISERRDMVADFQNRNDI
 FVFLSTRAGGLGINLTAXDTVHF (SEQ ID NO:671); TRMIDLLEEYMVYRK
 HTYXRLDGSSKISERRDM (SEQ ID NO:674); RRD MVADFQNRNDIFVFL

STRAGGLGINLTAXDTVHF (SEQ ID NO:675) , IFYDSDWNPTVDQQAMD
 RAHRLGQTKQVTYRLICKG (SEQ ID NO:676); RLICKGTIEERILQRAK
 EKSEIQRMVISG (SEQ ID NO:678). An additional embodiment is the polynucleotide
 fragments encoding these polypeptide fragments.

5 This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases and disorders of the brain. Similarly, polypeptides and
 10 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the central nervous system, expression of
 this gene at significantly higher or lower levels may be routinely detected in certain
 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
 15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
 individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder.

The tissue distribution and homology to a DNA helicase indicates that
 20 polynucleotides and polypeptides corresponding to this gene are useful for diseases
 affecting RNA transcription, particularly developmental disorders and healing wounds
 since the later are though to approximate developmental transcriptional regulation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 131

25 This gene is expressed primarily in prostate and to a lesser extent in amygdala
 and pancreatic tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 30 not limited to, prostate enlargement and gastrointestinal disorders, particularly of the
 pancreas and gall bladder. Similarly, polypeptides and antibodies directed to these
 polypeptides are useful in providing immunological probes for differential identification
 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
 particularly of the reproductive system, expression of this gene at significantly higher or
 lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
 another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of prostate diseases, including benign prostatic hyperplasia and prostate cancer. In addition, the tissue distribution in tumors of the pancreas indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in adult lung and to a lesser extent in hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary diseases and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pulmonary and respiratory disorders such as emphysema, pneumonia, and pulmonary edema and emboli. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 133

5 This gene is expressed primarily in human liver.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cirrhosis of the liver and other hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

 This gene is expressed primarily in fetal kidney and to a lesser extent in fetal liver and spleen.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development and regeneration of liver and kidney and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 367 as residues: Pro-70 to Arg-77, Tyr-102 to Thr-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the kidney and liver, such as cirrhosis, kidney failure, kidney stones, and liver failure, hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 135**

This gene is expressed primarily in brain, bone marrow, and to a lesser extent in placenta, T cell, testis and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and immunological diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 368 as residues: Met-1 to His-6.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also

play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 136

- 5 Translation product of this gene is homologous to the human WD repeat protein HAN11. Preferred polypeptide fragments comprise the following amino acid sequence:
- MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLLALGSFVEEYNNKVQLVG
LDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDYLRVWRVGETET
10 RLECLLNNNKNNSDFCAPLTSFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRV
NLVSGHVKTQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEH
STIIYEDPQHHPHLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTIE
HVSMAILGPHIIPATSALQRM TTRLSSGTSSKCPEPLRTL SWPTQLXGEINNVQ
WASTQPELSPSATTTAWRYSECSVG GAVPTRQGLLYFLPLPHPQS (SEQ ID
15 NO:679); MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLLALGSFV
EEYNNKVQLVGLDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDY
LRVWRVGETETRLECLLNNNKNNSDFCAPLTSFDWNEVDPYLL (SEQ ID
NO:680); SFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRVNLVSGHVK
TQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEHSTIIYEDPQH
20 HPI.LRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTI (SEQ ID
NO:681); VGADGSVRMFDLRHLEHSTIIYEDPQHHPHLLRLCWNKQDPNYLA
TMAMDGMEVVILDVRVPAHLXPGTTIEHVSMAILGPHIIPATSALQRM TTRL
SGTSSKCPEPLRTL SWPTQLXGEINNVQWASTQPELSPSATTTAWRYSECSVG
GAVPTRQGLLYFLPLPHPQS (SEQ ID NO:682). Also preferred are polynucleotide
25 fragments encoding these polypeptide fragments.

This gene is expressed primarily in placenta, embryo, T cell and fetal lung and to a lesser extent in endothelial, tonsil and bone marrow.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and developmental diseases in addition to cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above type, the polynucleotide or

polypeptide of the invention may be used for the detection of the above disorders in fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 369 as residues: Gly-19 to Gln-28, Pro-36 to Phe-42.

5 The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above
10 listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

 This gene is expressed primarily in TNF and INF induced epithelial cells, T cells and kidney.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory conditions particularly inflammatory reactions in the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of renal
25 system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 370 as residues: Thr-67 to Gly-72, Gln-132 to Ala-145, Arg-150 to Pro-157.

 The tissue distribution indicates that the protein products of this gene are useful for treating the damage caused by inflammation of the kidney.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. D63485).

This gene is expressed primarily in breast cancer and colon cancer and to a lesser extent in thymus and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers, especially of the breast and colon tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene maps to chromosome 17, and therefore, can be used as a marker for linkage analysis from chromosome 17.

This gene is expressed primarily in CD34 positive cells, and to lesser extent in activated T-cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunologically related diseases and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoietic system, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34, T-cell and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of hematopoietic disorders and immunologically related diseases, such as anemia, leukemia, inflammation, infection, allergy, immunodeficiency disorders, arthritis, asthma, immune deficiency diseases such as AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene was recently cloned by another group, who called the gene KIAA0313 gene. (See Accession No. d1021609.) Preferred polypeptide fragments comprise the amino acid sequence:

LYATATVISSPSTEXLSQDQGDRA SLDAADSGRGSWTSCSSGSHDNIQTIQ
HQRSWETLPFGHTHFDYSGDPAGLWASSSHMDQIMFSDHSTKYNRQNSRES
LEQAQSRASWASSTGYWGEDSEGDGTIKRRGGKDV SIEAESSSLTSVTTEETK
PVPMPAHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITDFPEGHSH PARKP
PDYNVALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQWHKXNESDPR
LAPYQSQGFSTEEDEDEQVSAV (SEQ ID NO:683); HMDQIMFSDHSTKYNRQ
NQSRESLEQAQSRASWASSTGYWGE (SEQ ID NO:684); SVTTEETKPVPMP
AHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITD (SEQ ID NO:685); and
VALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQW

HKXNESDPR LAPYQSQGF (SEQ ID NO:686). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 4, and therefore, may be used as a marker in linkage analysis for chromosome 4 (See Accession No. AB002311).

This gene is expressed primarily in ovarian cancer, tumors of the Testis, brain, and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, ovarian, testicle, brain and colon cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male and female reproductive systems.

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, testis, and brain origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in spleen and colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, colon cancer and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal tract and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 142

Translation product is homologous to T cell translocation protein, a putative zinc finger factor (See Accession No. 340454), as well as to the G-protein coupled receptor TM5 consensus polypeptide (See Accession No. R50734). Preferred polypeptide

5 fragments comprise the following amino acid sequence:

CLLFVVFVSLGMRCLEWTIVYNVLYLKHKCNVLLCYHLCSI (SEQ ID NO:687);

ACSKLIPAFEMVMRAKDNVYHLDCFACQLCNQRXCVGDKFFLKNNXXLCQT

DYEEGLMKEGYAPXVR (SEQ ID NO:688). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

10 This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders including brain cancer. Similarly, polypeptides

15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central Nervous System,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,

20 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

25 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with

30 the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 143

Translation product for this gene has significant homology to the Fas ligand, which is a cysteine-rich type II transmembrane protein/tumor necrosis factor receptor

35 homolog. Mutations within this protein have been shown to result in generalized lymphoproliferative disease leading to the development of lymphadenopathy and autoimmune disease (See Medline Article No. 94185175). Preferred polypeptide

fragments comprise the following amino acid sequence:

SALSEPGAPDRRRPCPESVPRRPDDEQWPPPTALCLDVAPLPPSS (SEQ ID NO:689). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. 473565).

5 This gene is expressed primarily in osteoblasts, lung, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoblast-related, pulmonary, neurological, and immunological
10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
15 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 376 as residues: Trp-33 to Thr-40, Lys-
20 45 to Ile-63.

The tissue distribution in osteoblasts, lung, and brain combined with its homology to the Fas ligand indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as,
25 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the Fas ligand gene is known to be expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including asthma, immune deficiency diseases such as AIDS
30 and leukemia, and various autoimmune disorders including lupus and arthritis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene shares sequence homology with a 21.5 KD transmembrane protein in the SEC15-SAP4 intergenic region of yeast (See Accession No. U72207). The

...KSTLAKRSKSCUTDGFQGLVANKGRREKAWAHL
SGPGGGSRGSRSDRGSGQGDSLYPVGYLDKQVPDTSVQETDRIILVEKRCWDIAL

GPLKQIPMNLFIMYMAGNTISIFPTMMVCMMAWRPIQALMAISATFKMLESSSQ
 KFLQGLVYLIGNLMGLALAVYKCQSMGLLPTHASDWLAFIEPPERMEFSGG
 GLLL (SEQ ID NO:691); PVGYLDKQVPDTSVQETDRILVEKRCWDIALGPLKQ
 IPMNLFI (SEQ ID NO:693); and ATFKMLESSSQKFLQGLVYLIGNLMGLALAV
 5 YKCQSMGLLPTHASD (SEQ ID NO:692). Also preferred are polynucleotide
 fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma, hemangiopericytoma, liver,
 lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
 10 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, osteoclastoma, hemangiopericytoma, liver and lung tumors. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 immunological probes for differential identification of the above tissue(s) or cell
 15 type(s). For a number of disorders of the above tissues or cells, particularly of the lung
 and liver systems, expression of this gene at significantly higher or lower levels may be
 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
 cell sample taken from an individual having such a disorder, relative to the standard
 20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosing osteoclastoma,
 hemangiopericytoma, liver and lung tumors.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 145

Translation product of this gene shares homology with the glucagon-69 gene
 which may indicate this gene plays a role in regulating metabolism. (See Accession No.
 A60318) One embodiment for this gene is the polypeptide fragments comprising the
 30 following amino acid sequence:
 PTTKLDIMEKKKHHIQIRFPSFYHKLVDSGRMRSKRETRREDSDTKHNL (SEQ ID
 NO:694). An additional embodiment is the polynucleotide fragments encoding these
 polypeptide fragments.

This gene is expressed primarily in brain, kidney, colon, and testis.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, brain, kidney, colon, and testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, neurological, circulatory, and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of brain, kidney, colon, and testis origins, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 146**

The translation product of this gene shares sequence homology with goliath protein which is thought to be important in the regulation of gene expression during development. Protein may serve as a transcription factor. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

30 TEHIIAVMITELRGKDILSYLEKNISVQMTIavgTRMPpKNFSRGSLSVFVSISFIV
LMISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKETD
PDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNILKA
LGIV (SEQ ID NO:695); TEHIIAVMITELRGKDILSYLEKNISVQMTIavgTRMP
PKNFSRGSLSVFVSISFIV LMISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKETD

PDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGIV (SEQ ID NO:696);

WLSEHCTCPMCKLNILKALGIV (SEQ ID NO:699). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. 157535). Moreover, another embodiment is the polynucleotide fragments encoding these polypeptide fragments:

- 5 MTHPGTEHIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGS
LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTV
KKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCP
MCKLNILKALGIVPNLPCTDNVAFDMERLTRTQAVNRRSALGDLAGDNSLGLE
PLRTSGISPLPQDGELTPRTGEINIAVTKEWFIIASFGLLSALTLCYMIIRATASLN
10 ANEVEWF (SEQ ID NO:696); MTHPGTEHIIAVMITELRGKDILSYLEKNISVQM
TIAVGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR
LGDAAKKAISKLTTRT (SEQ ID NO:700); AAKKAISKLTTRTVKKGDKET
TDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNIL
KALGIVPNLPC (SEQ ID NO:701); TQAVNRRSALGDLAGDNSLGLEPLRTSGI
15 SPLPQDGELTPRTGEINIAVTKEWFIIASFGLLSALTLCYMIIRATASLNANEVEW
F (SEQ ID NO:702); PLHGVADHLGCDPQTRFFVPPNIKQWIALLRGNCTF
KEKISRAAFHNAVAVVIYNNKSKEEPVTMTHPGTEHIIAVMITELRGKDILSYLE
KNISVQMTIAVGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTN
RDRNQRR LGDAAKKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVR
20 LPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGIVPNLPCTDNVAFDMERLT
RTQAVNRRSALGDLAGDNSLGLEPLRTSGISPLPQDGELTPRTGEINIAVTKEW
FIIASFGLLSALTLCYMIIRATASLNANEVEWF (SEQ ID NO:703); and
HGVADHLGCDPQTRFFVPPNIKQWIALLRGNCTFKEKISRAAFHNAVAVVIY
NNKSKEE (SEQ ID NO:704). An additional embodiment is the polynucleotide
25 fragments encoding these polypeptide fragments. When tested against Jurkat cell lines,
supernatants removed from cells containing this gene activated the GAS pathway.
Thus, it is likely that this gene activates immune cells through the JAKS/STAT signal
transduction pathway.

- 30 This gene is expressed primarily in macrophage, breast, kidney and to a lesser
extent in synovium, hypothalamus and rhabdomyosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, schizophrenia and cancer. Similarly, polypeptides and antibodies directed
35 to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune and neural system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level
 5 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zinc finger protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of schizophrenia, kidney disease and other cancers. The tissue distribution in macrophage, breast, and kidney origins indicates that polynucleotides and polypeptides
 10 corresponding to this gene are useful for diagnosis and intervention of tumors within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in
 15 immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 147

20 The translation product of this gene shares sequence homology with HNP36 protein, an equilibrative nucleoside transporter, which is thought to be important in gene transcription as well as serving as an important component of the nucleoside transport apparatus (See Accession No. 1845345). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:
 25 MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTHICYLGLPRLEFYR
 YYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGVSVSNSQPTNESHSHKAILK
 NISVLAFSVCFITITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTENIFDWLG
 RSLTAVFMWPGKDSRWLPSWXLARLVFVPLLLLCNIKPRRYLTVVFEHDAWFI
 FFMAAFASFNGYLASLCMCFGPKKVKPAEAEAEPSWPSSCVVWHWGLFS
 30 PSCSGQLCDKGWTEGLPASLPVCLLPARGDPEWSGGFFF (SEQ ID
 NO:705); MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTHIC
 YLGLPRLEFYRYYQQLKLE GPGEQETKLDLISKGEEPRAGKEESGVSVSNSQ
 PTNESHSHI (SEQ ID NO:706); SGVSVSNSQPTNESHSHKAILKNISVLAFSVCFI
 FTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTENIFDWLGRS (SEQ ID

NO:708); FGPKKVKPAEAEAEPSWPSSCVVWHWGLFSPSCSGQLCDK

GWTEGLPASLPVCLLPSPARGDPEWSGGFFF (SEQ ID NO:709). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in eosinophils and aortic endothelium and to a lesser extent in umbilical vein endothelial cell and thymus.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to HNP36 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of blood neoplasias and other hematopoietic disease.

20 **FEATURES OF PROTEIN ENCODED BY GENE NO: 148**

This gene is expressed primarily in breast cancer cell lines, thymus stromal cells, and ovary.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and female reproductive system diseases including breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
30 type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
35 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endocrine disorders. In addition, the tissue distribution in tumors of thymus, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 149**

Translation product of this gene has homology to pmt1 and pmt 2, two conserved schizosaccharomyces pombe genes. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:
 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTFNEDEG
 15 ELPEWVFVQEEKQHRIRQLPVGKKEVEHYRKRWREINARPIXXXXXXXXXXXXX
 XXXXXXLEQTRKKAEAVVNTVDIXRTRES (SEQ ID NO:710);
 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTF (SEQ
 ID NO:711); KRWREINARPIXXXXXXXXXXXXXXXXXXXXXLEQTRKKAE
 AVVNTVDIXRTRES (SEQ ID NO:712). An additional embodiment is the
 20 polynucleotide fragments encoding these polypeptide fragments (See Accession No. e1216734).

This gene is expressed primarily in retina and ovary and to a lesser extent in breast cancer cell, epididymus and osteosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal growth disorders, cancer and reproductive system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
 30 type(s). For a number of disorders of the above tissues or cells, particularly of the neural and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having a disorder of the above tissues or cells.

Polynucleotides and polypeptides of the invention include those comprising a sequence shown in SEQ ID NO: 382 as residues: Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of reproductive system disease and cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 150

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKEKKRNKKKKKTIGSPKRIQS
PLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLS

10 SLQSDPAGCVRPPAPNLAGAVEFNDVKTLREWITTISDPMEEDILQVVKYCTD
LIEEKDLEKLDLVIKYMKRLMQQSVEVWNMAFDNFILDNVQVVLQQTYGSTLK

VT (SEQ ID NO:713); MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKE

KKRNKKKKKTIGSPKRIQ (SEQ ID NO:714); KRIQSPLNNKLLNSPAKT

LPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPP

15 APNLAGAVEFNDVKTLREWITTISDPM (SEQ ID NO:715);

TISDPMEEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVE

SVWNMAFDNFILDNVQVVLQQTYGSTLKVT (SEQ ID NO:716). An additional

embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in 12 week embryo and to a lesser extent in
20 hemangiopericytoma and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, growth disorders and hemangiopericytoma. Similarly, polypeptides and
25 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the circular and neural system, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
NO: 383 as residues: Leu-4 to Lys-11.

35 The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the treatment of growth disorders,
hemangiopericytoma and other soft tissue tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 151

The translation product of this gene has been found to have homology to a human DNA mismatch repair protein PMS3. Preferred polypeptide fragments comprise the following amino acid sequence: FCHDCKFPEASPAMNCEP (SEQ ID NO:717). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R95250).

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphoma, immunodeficiency diseases, and cancers resulting from genetic instability. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 384 as residues: Met-1 to Lys-6.

The tissue distribution in neutrophils and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Furthermore, its homology to a known DNA repair protein would suggest gene may be useful in establishing cancer predisposition and prevention in gene therapy applications.

FEATURES OF PROTEIN ENCODED BY GENE NO: 152

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious diseases and lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of inflammation and infectious diseases.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 153**

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLFPVLRKKC
 NFFCWDSSAHSPLHPLSASCSAPACHASDTHLLYPSTRALCPSIFAWLVAPHS
 20 VERTNAPGPTPSSQSSPVFPVFPVSFEMALIVCXLVCC (SEQ ID NO:720);
 MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLFPVLRKKCNFFCWDSSAH
 SLPLHPLSASCSAPACHA (SEQ ID NO:721);FAWLVAPHSVERTNAPGPTPS
 SQSSPVFPVFPVSFEMALIVCXLVCC (SEQ ID NO:722). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 386 as residues: Ser-11 to Pro-17.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases and inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed in multiple tissues including ovary, uterus, adipose tissue, brain, and the liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, uterine, ovarian, brain, and liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in the treatment of the female reproductive system, obesity, and liver disorders, particularly cancer in the above tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 155

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. D87452).

This gene is expressed in multiple tissues including brain, aortic endothelial cells, smooth muscle, pituitary, testis, melanocytes, spleen, neutrophils, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in the treatment of various disorders, including immunodeficiencies, cancers, and disorders of the brain and the female reproductive system, as well as cardiovascular disorders, such as

atherosclerosis and stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution suggest that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nervous system, including schizophrenia, neurodegeneration, neoplasia, brain cancer as well as cardiovascular and female reproductive disorders including cancer within the above tissues.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 156**

The translation product of this gene shares sequence homology with the human gene encoding cytochrome b561 (See Accession No. P10897). Cytochrome b561 is a transmembrane electron transport protein that is specific to a subset of secretory vesicles

20 containing catecholamines and amidated peptides. This protein is thought to supply reducing equivalents to the intravesicular enzymes dopamine-beta-hydroxylase and alpha-peptide amidase. Preferred polypeptides of the invention comprise the amino acid sequence:

MAMEGYWRFLALLGSALLVGFLSVIFALVWVLHYREGLGWDGSALEFNWHP
 25 VLMVTGFVFIQGIHIVYRLPWTWKCSKLLMKSIHAGLNAVAAILAISVVAVFE
 NHNVNNIANMYSLHSWVGLIAVICYLLQLLSGFSVFLLPWAPLSLRAFLMPIHV
 YSGIVIFGTVIATALMGLTEKLIFSLRDPAYSTFPPEGVFVNTLGLLLLVFGALIF
 WIVTRPQWKRPKEPNSTILHPNGGTEQGARGSMPAYSGNNMDKSDSEL

30 NSEVAARKRNALDEAGQRSTM (SEQ ID NO:724); as well as antigenic fragments of at least 20 amino acids of this gene and/or biologically active fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system and metabolism related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product or RNA of this gene is useful for treatment or diagnosis of immune system and metabolic diseases or conditions including Tay-Sachs disease, phenylketonuria, galactosemia, various porphyrias, and Hurler's syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 157

The translation product of this gene shares sequence homology with collagen which is important in mammalian development. This gene also shows sequence homology with bcl-2. (See Accession No. P80988.) Preferred polypeptide fragments comprise the amino acid sequence: PGRAGPSPGLSLQLPAEPGHPAGNLAPLTSRPQPLCRIPAVPG (SEQ ID NO:725). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

This gene is expressed primarily in HL-60 tissue culture cells and to a lesser extent in liver, breast, and uterus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases, hereditary disorders involving the MHC class of immune molecules, as well as developmental disorders and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having the disorder.

It is to be understood that the invention is not limited to the specific embodiments disclosed herein, but that it includes all equivalents thereof within the scope of the claims.

comprising a sequence shown in SEQ ID NO: 390 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62.

The tissue distribution and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hereditary MHC disorders and particularly autoimmune disorders including rheumatoid arthritis, lupus, scleroderma, and dermatomyositis, as well as many reproductive disorders, including cancer of the uterus, and breast tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 158

This gene is expressed primarily in the amygdala region of the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, particularly those effecting mood and personality. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and/or diagnosis of a variety of brain disorders, particularly bipolar disorder, unipolar depression, and dementia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 159

This gene is expressed in a variety of tissues and cell types including brain, smooth muscle, kidney, salivary gland and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers of a variety of organs including brain, smooth muscle, kidney, salivary gland and T-cells and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the central nervous, urinary, salivary, digestive, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, smooth muscle, and T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of various neurological, and cardiovascular disorders, but not limited to cancer within the above tissues. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with collagen which is thought to be important in cellular interactions, extracellular matrix formation, and has been found to be an identifying determinant in autoimmune disorders. Moreover, this gene shows sequence homology with the yeast protein, Sls1p, an endoplasmic reticulum component, involved in the protein translocation process in Yeast *Yarrowia lipolytica*. (See Accession No. 1052828; see also J. Biol. Chem. 271, 11668-11675 (1996).) With mouse, this same region shows sequence homology with the heavy chain of kinesin. (See Accession No. 2062607.) Recently, suppression of the heavy chain of kinesin was shown to inhibit insulin secretion from primary cultures of mouse beta-cells. (See Endocrinology 138 (5), 1979-1987 (1997).) Moreover, kinesin was found associated with drug resistance and cell immortalization. (See 468355.) Thus, it is likely that this gene also act as a genetic suppressor elements.

This gene is expressed primarily in the greater omentum and to a lesser extent in a variety of organs and cell types including gall bladder, stromal bone marrow cells, lymph node, liver, testes, pituitary, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissues and cell types.

These polynucleotides and polypeptides are useful for diagnosis of various disorders, including autoimmune disorders and cancers in a variety of organs and cell types.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 393 as residues: Asn-27 to Leu-47, Gln-81 to Lys-88, Asp-93 to Lys-102, Asn-107 to Leu-116, Met-129 to Glu-141, Glu-150 to Asp-157, Lys-176 to Glu-185, Glu-333 to Tyr-349, Cys-393 to Leu-403, Gln-423 to Gly-429.

The tissue distribution in within various endocrine and immunological tissues combined with the sequence homology to a conserved collagen motif indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various autoimmune disorders including, but not limited to, rheumatoid arthritis, lupus erythematosus, scleroderma, dermatomyositis. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene has homology to the tissue inhibitor of metalloproteinase 2. Such inhibitors are vital to proper regulation of metalloproteins such as collagenases (See Accession No. P16368). In addition, this gene maps to chromosome 17, and therefore, may be used as a marker in linkage analysis for chromosome 17 (See Accession No. P16368).

This gene is expressed primarily in several types of cancer including osteoclastoma, chondrosarcoma, and rhabdomyosarcoma and to a lesser extent in several non-malignant tissues including synovium, amygdala, testes, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, various types of cancer, particularly cancers of bone and cartilage, as well as various autoimmune disorders. Similarly, polypeptides and antibodies directed

to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various cancers and the sequence homology to a collagenase inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 162

This gene is homologous to the mitochondrial ATP6 gene and therefore is likely a homolog of this gene family (See Accession No. X76197).

This gene is expressed primarily in brain tissue.

20

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, including Down's syndrome, depression, Schizophrenia, and epilepsy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

30

The tissue distribution in brain tissue indicates this gene is useful for diagnosis of various neurological disorders including, but not limited to, Down's syndrome, depression, Schizophrenia, and epilepsy.

EXAMPLE 1: SCREENING FOR THE APPLICATION OF THE INVENTION TO THE DETECTION OF DOWN'S SYNDROME, depression, Schizophrenia, and epilepsy, phenylketonuria and Hurler's Syndrome

FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in placenta, neutrophils, and microvascular endothelial cells and to a lesser extent in multiple tissues including brain, prostate, spleen, thymus, and bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neutropenia and other diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis various female reproductive disorders. Additionally the gene product may be used as a target in the immunotherapy of various cancers. Because the gene is expressed in some cells of lymphoid and endocrine origin, the natural gene product may be involved in immune functions and metabolism regulation, respectively. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 164

This gene is expressed primarily in neutrophils, monocytes, bone marrow, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders including, but not limited to, autoimmune disorders such as lupus, and immunodeficiency disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 5 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various immune system tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various immunological disorders such as Hodgkin's lymphoma, arthritis,
 10 asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 165

The translation product of this gene shares sequence homology with dystrophin which is thought to be defective in both Duchene and Becker Muscular Dystrophy.
 15 Preferred polypeptide fragments comprise the following amino acid sequence:
 MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDRWELLQAQ
 ALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELSTDIQTIELQ
 IKKLKELQKAVDHRKAILLSINLCSPEFTQADSKESRDLQDRLXQMNGRWDRV
 CSLLEEWGRLQDALMQCQGFHEMSHGLLLMLENIDRRKNEIVPIDSNLDAEIL
 20 QDHHKQLMQIKHELLESQLRVASLQDMSCQLLVNAEGTDCLEAKEKVHVIGNR
 LKLLLKEVSRHIKELEKLLDVSSSQQDI.SSWSSADELDTSGSVSPXSGRSTPNR
 QKTPRGKCSLSQGPSVSSPHSRSTKGGSDSSLSEXPGRSGRGFLFRVLRAA
 LPLQLLLLLLIGLACLVPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:726); MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDR
 25 WELLQAQALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELS
 TDIQTIELQIK (SEQ ID NO:727); KLKELQKAVDHRKAILLSINLCSPEFTQADSK
 ESRDLQDRLXQMNGRWDRVCSLLEEWGRLQDALMQCQGFHEMSHGLLLML
 ENIDRRKNEIVPIDSNLDAEILQDHHKQLMQIKHELLESQLRVASLQDMSCQL
 (SEQ ID NO:728); QDMSCQLLVNAEGTDCLEAKEKVHVIGNRLKLLLKEVS
 30 RHIKELEKLLDVSSSQQDI.SSWSSADELDTSGSVSPXSGRSTPNRQKTPRGKCS
 LSQGPSVSSPHS (SEQ ID NO:729); DSSLSEXPGRSGRGFLFRVLRAAL
 PLQLLLLLLIGLACLVPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:730). Also preferred are polynucleotide fragments encoding these polypeptide
 fragments. Furthermore, this gene may be used to identify and isolate cDNA clones

EXAMPLE 1: EXPRESSION OF HUMAN CD34 mRNA IN THE TISSUES OF HUMAN BRAIN

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, musculoskeletal disorders including Muscular Dystrophy and cardiovascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscle tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to dystrophin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of Muscular Dystrophy and other muscle disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the central nervous system, including Alzheimer's Disease, Parkinson's Disease, ALS, and mental illnesses. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 399 as residues: Pro-20 to Gly-26, Leu-37 to Pro-42, His-57 to Gly-63.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the central nervous system and may protect or

enhance survival of neuronal cells by slowing progression of neurodegenerative diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 167

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MKLLICGNYLAPSHSESSRRCCLLCFYPLCLEINFGMKVFLSMPFLVLFQ

10 SLIQED (SEQ ID NO:731). Polynucleotides encoding such polypeptides are also provided. This gene is believed to reside on chromosome 15. Therefore polynucleotides derived from this gene are useful in linkage analysis as chromosome 15 markers.

This gene is expressed primarily in human testes tumor and to a lesser extent in normal human testes.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the testes, particularly cancer, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of 20 the male reproductive tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily 25 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of testicular diseases including cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 168

30 This gene is expressed primarily in fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, conditions affecting hepatocellular carcinoma, leukemia, and lymphoma.

Similarly, immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

hepatic system, and fetal hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 401 as residues: His-7 to Trp-17, Leu-19 to Lys-27, Pro-33 to Gly-44, Lys-68 to Gly-74, Lys-85 to Cys-95.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the developing liver and hematopoietic system, and act as a growth differentiation factor for hematopoietic stem cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The polypeptide encoded by this gene is believed to be a membrane bound receptor. The extracellular domain of which is expected to consist of the following amino acid sequence:

RILLVKYSANEENKYDYLPPTTVNVCSELVKLVFCVLVSFCVIKKDHQSRNLKY
 ASWKEFSDFMKWSIPAFLYFLDNLIVFYVLSYLQPAMAVIFSNSFSIITTALLFRIV
 LKXRLNWIQWASLLTLFLSIVALTAGTKTLQHNLAGRGFHHDAFFSPSNSCLL
 FRNECPRKDNCTAKEWTFPEAKWNTTARVFSHIRLGMGHVLIIVQCFISSMANI
 YNEKILKEGNQLTEXIFIQNSKLYFFGILFNGLTLGLQRSNRDQIKNCGFFYGH
 S (SEQ ID NO:732). Thus, preferred polypeptides encoded by this gene comprise the extracellular domain as shown above. It will be recognized, however, that deletions of either end of the extracellular domain up to the first cysteine from the N-terminus and the first cysteine of the C-terminus, is expected to retain the biological functions of the full-length extracellular domain because the cysteines are thought to be responsible for providing secondary structure to the molecule. Thus, deletions of one or more amino acids from either end (or both ends) of the extracellular domain are contemplated. Of course, further deletions including the cysteines are also contemplated as useful as such polypeptides is expected to have immunological properties such as the ability to evoke and immune response. Polynucleotides encoding all of the foregoing polypeptides are provided.

This gene is expressed primarily in human osteoclastoma and to a lesser extent in hippocampus and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancers, particularly those of the bone and connective tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of cancers of the bone and connective tissues, and may act as growth factors for cells involved in bone or connective tissue growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 170

Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

NSVPNLQTLAVLTEAIGPEPAIPRXPREPPVATSTPATPSAGPQPLPTGTV

LVPGGPAPPCLGEAWALLLPPCRPSLTSCFWSPRPSPWKETGV (SEQ ID NO:733). Polynucleotides encoding such polypeptides are also provided herein.

This gene is expressed primarily in hematopoietic progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the blood including cancer and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the blood/circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84, His-101 to His-109, and His-32.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of diseases involving growth differentiation of hematopoietic cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 171

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequences: ALQLAFYPDAVEEWLEENVHPSLQRLQXLLQDLSEVSAPP (SEQ ID NO:734); and/or CHPPALAGTLLRTPEGRAHARGLLLEAGGA (SEQ ID NO:735). Polynucleotides encoding such polypeptides are also provided. The protein product of this gene shares sequence homology with metallothionines. Thus, polypeptide encoded
10 by this gene are expected to have metallothionine activity, such activities are known in the art and described elsewhere herein.

This gene is expressed primarily in kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the kidney including cancer and renal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system,
20 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
25 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 404 as residues: Ser-47 to Gln-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the kidney including kidney failure.
30

FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 405 as residues: Gln-31 to Thr-43, Gly-51 to Ser-58, Pro-65 to Pro-72.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of developmental problems with fetal tissue. The gene may be involved in vital organ development in the early stage, especially hematopoiesis, cardiovascular system, and neural development.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 173**

The translation product of this gene shares sequence homology with TGN38, an integral membrane protein previously shown to be predominantly localized to the trans-Golgi network (TGN) of cells.

This gene is expressed primarily in developing embryo and to a lesser extent in cancer tissues including lymphoma, endometrial, prostate and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 406 as residues: His-65 to Ser-72, Pro-82 to Gly-90, Pro-96 to Gly-103, and Pro-109 to Ser-116.

The tissue distribution and homology to an integral membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for

diagnosis of cancers and developmental abnormalities where aberrant expression relates to an abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 174

5 The translation product of this gene shares sequence homology with a dnaJ heat shock protein from *E. coli* which is allelic to sec63, a gene that affects transit of nascent secretory proteins across the endoplasmic reticulum in yeast.

 This gene is expressed primarily in Hodgkin's lymphoma and to a lesser extent in testes.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to
20 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 407 as residues: Thr-13 to Trp-21, Arg-74 to Asp-81.

 The tissue distribution and homology to dnaJ indicates that polynucleotides and
25 polypeptides corresponding to this gene are useful as a diagnostic for cancer including Hodgkin's lymphoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 175

 This gene is expressed primarily in endothelial cells and to a lesser extent in
30 bone marrow stromal cells.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving angiogenic abnormalities including diabetic
35 retinopathy, macular degeneration, and other diseases including arteriosclerosis and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treating diseases where an increase or decrease in angiogenesis is indicated and as a factor in the wound healing process.

FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with MAT8 (mouse) which is thought to be important in regulating chloride conductance in cells (particularly in the breast) by modulating the response mediated by cAMP and protein kinase C to extracellular signals.

This gene is expressed primarily in amniotic cells and hematopoietic cells including macrophages, Neutrophils, T cells, TNF induced aortic endothelium and to a lesser extent in testes, TNF induced epithelial cells, and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory responses mediated by T cells, macrophages, and/or neutrophils particularly those involving TNF, and also cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 409 as residues: Thr 19 to Ala 33, Leu 51 to Arg 82, Pro 90 to Ser 107, Pro 114 to

Due to its homology to MAT8, it is indicated that polynucleotides and polypeptides corresponding to this gene are useful for modifying inflammatory

responses to cytokines such as TNF and thus modifying the duration and/or severity of inflammation. Polynucleotides and polypeptides derived from this gene are thought to be useful in the diagnosis and treatment of cancer.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vascular restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases associated with vascular response to injury such as vascular restenosis following angioplasty..

FEATURES OF PROTEIN ENCODED BY GENE NO: 178

One embodiment of the claimed invention comprises:

MRPDWKAGAGPGGPPQKPAPSSQRKPPARPSAAAAIAVAAAEERRLRQRN
RLRLEEDKPAVERCLEELVFGDVENDEALLRRLRGPRVQEHEDSGDSEVENEAKGNFPPQKKPV
WVDEEDEDEEMVDMMNRRFRKDMMNASESKLSKDNLKKRLKEEFQHAMGGVPAWAETTKRKTSSDDESEDEDDLLQRTGNFISTSTSLPRG
ILKMKNCQHANAERPTVARISICAVPSRCTDCDGCWD (SEQ ID NO:737); or
CLEELVFGDVENDEALLRRLRGPRVQEHEDSGDSEVENEAKGNFPPQKKPV
WVDEEDEDEEMVDMMNRRFRKDMMNASESKLSKDNLKKRLKEEFQHAMG
GVPWAETTKRKTSSDDESEDEDDLLQRTGNFISTSTSLPRGILKMKNCQHA
NAERPTVARISICAVPSRCTDCDGC (SEQ ID NO: 738). LKEKIVRSFEVSPDGS
FLINGIAGYLHLLAMKTKELIGSMKINGRVAASTFSSDSKKVYASSGDGEVYV
WDVNSRKCLNRFVDEGSYGLSIATSRNGQYVACGSNCGVVNIYNQDSCLQE
TNPKPIKAIMNLVTGVTSLTFNPTTEILAIASEKMKEAVRLVHLPSTVFSNFPVI
KNKNISHVHTMDFSPRSGYFALGNEKGKALMYRLHHYSDF (SEQ ID NO:739);

and/or KINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSL
YGLSIATSRNGQYVACGSNCGVVNIYNQDSCLQETNPKPIKAIMNLVTGVTSLT
FNPTTEILAIASEKMKEAVRLVHLPSTVFSNFPVIKKNKISHVHTMDFSPRSG
YFALGNEKGKAL (SEQ ID NO:740).

- 5 This gene is expressed primarily in epididymus and endometrial tumors and to a lesser extent in T cell lymphoma and cell lines derived from colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of the reproductive organs including testis and endometrial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 411 as residues: Ser-67 to Lys-72, Val-87 to Leu-93, Tyr-128 to Pro-141, Asp-204 to Gly-210.

The tissue distribution indicates that the protein products of this gene are useful for treating tumors of the endometrium or epithelial tumors of the reproductive system.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 179**

Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MRILQLILLALATGLVGGETRIIKGFECKLHSQPWQAALFEKTRLLCGATLIAPR
WLLTAAHCLKPRYIVHLGQHNLQKEEGCFQTRTATESFPHIPGFNNSLPNKDH
30 RNDIMLVKMASPV SITWAVRPLTLSSRCVTAGTSCSFPAGAARPDPSYACLTPC
DAPTSPSLSTRSVRTPTPATSQTPWCVPACRKGARTPARVTPGALWSVTSIFKA
LSPGARIRVRSPESLVSTRKSANMWTGSRRR (SEQ ID NO:741); ETRIIKGFEC
KLHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNLQKEE
GCFQTRTATESFPHIPGFNNSLPNKDH RNDIMLVKMASPV SITWAVRPLTLSSRCVTAGTSCSFPAGAARPDPSYACLTPC

WLLTAAHCLKPRYIVHLGQHNLQKEEGCFQTRTATESFPHIPGFNNSLPNKDH RNDIMLVKMASPV SITWAVRPLTLSSRCVTAGTSCSFPAGAARPDPSYACLTPC

SRRR (SEQ ID NO:742); or CKLHSQPWQAALFEKTRLLCGATLIAPRWLLT
AAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNS

(SEQ ID NO:743). The translation product of this gene shares sequence homology
with neuropsin a novel serine protease which is thought to be important in modulating
extracellular signaling pathways in the brain. Owing to the structural similarity to other
serine proteases the protein products of this gene are expected to have serine protease
activity which may be assayed by methods known in the art and described elsewhere
herein.

This gene is expressed primarily in endometrial tumor and to a lesser extent in
colon cancer, benign hypertrophic prostate, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cancers of the endometrium or colon and benign hypertrophy of the
prostate. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the tissue(s)
or cell type(s). For a number of disorders of the above tissues or cells, particularly of
the urogenital or reproductive systems, expression of this gene at significantly higher or
lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
comprising a sequence shown in SEQ ID NO: 412 as residues: Gly-12 to Ser-22, Pro-
34 to Ser-53.

The tissue distribution and homology to serine proteases indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diagnosing
or treating hyperproliferative disorders such as cancer of the endometrium or colon and
hyperplasia of the prostate.

FEATURES OF PROTEIN ENCODED BY GENE NO: 180

Preferred polypeptide encoded by this gene comprise the following amino acid
sequence: VLQGRYFSPILEMRRLRPEGXXNLPGGSSRAQKEPRQDLTLVLWPHC
PHFAMTRSYPVKQCMVQGSFYCIFYKGPVQNWNC (SEQ ID NO:744).

Polynucleotides encoding such polypeptide are also provided.

This gene is expressed primarily in fetal brain

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, identifying and expanding stem cells in the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for detecting and expanding stem cell populations in the (or of the) central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in early stage human brain and a stromal cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 414 as residues: Gln 12 to Gln 17, Gln 51 to Pro 60.

are useful for diagnosing or treating developmental abnormalities of the central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 182

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MPIIDQVNPPELHDFMQSAEVTIFALSWLITWFGHVLSDFRHVVRLYDF
 FLACHPLMPIYFAAVIVLYREQEVLDCDCDMASVHHLLSQIPQDLPYETLISRXE
 TFLFSFPHPNLLGRPLPNSKLRGRQPLLSKTLSTWHQPSRGLIWCCGSGXRGLL
 10 RPEDRTKDVLTKPRTNRFVKLAVMGLTVALGAAALAVVKSALAWAPKFQLQL
 FP (SEQ ID NO:745); or CPEFFIPATLPCPFVFAFTSEASSRAYLTQRGPGGLAQ
 NLMPLPVGFWMGSLPPPWCWRKWVSEACSCFC (SEQ ID NO:746) These
 polypeptides are structurally similar to various TGF-beta family members. Thus, this
 polypeptide is expected to have a variety of activities in the modulation of cell growth
 15 and proliferation.

This gene is expressed primarily in osteoclastoma, microvascular endothelium, and bone marrow derived cell lines.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological diseases particularly involving aberrant proliferation of stem cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
 25 the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 415 as residues: Ser-33 to Ala-39.

The tissue distribution indicates that the protein products of this gene is useful for treating disorders of the progenitors of the immune system. Applications include in vivo expansion of progenitor cells, ex vivo expansion of progenitor cells, or the
 35 treatment of tumors of the circulatory system, such as lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 183

This gene maps to chromosome 17 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 5 GFGSVSAAGRRSGGTWQPVQ (SEQ ID NO:747); PGGLAVGSRWWSRSLT (SEQ ID NO:748); LEPSRQRRPRRRGGTSRPETDQRAKCWRQL (SEQ ID NO:749); and/or VCLRCQNRMEN (SEQ ID NO:750). In further specific embodiments, polypeptides of the invention comprise the sequence: MAACTARRPGR GQPLVVPVADXGPVAKAALCAAXAGAFSPASTTTTTRHLSSNRNPEGKVLETV
- 10 GVFEVPKQNGKYETGQLFLHSIFGYRGVVLPWQARLXDRDVASAAPEKAEN PAGHGSKEVKGKTHTYQVLIDARDCPHISQRSQTEAVTFLANHDDSRALYAIP GLDYVSHEDILPYTSTDQVPIQHELPERFLLYDQTKAPPFVARETLRAWQEKNH PWLELSDVHRETTENIRVTVIPFYMGMRQAQNSHVYWWRYCIRLENLSDSDVVQ LRERHWRIFSLSGTLETVRGRGVVGREPVLSEKQPAFYSSHVSLQASSGHMW
- 15 GTFRFRPDGSHFDVRIPFSLESNKDEKTPPSGLHW (SEQ ID NO:751); MAACTARRPGRGQPLVVPVADXGPVAKAALCAA (SEQ ID NO:752); VLETVG VFEVPKQNGKYETGQLFLHSIFGYRGVVL (SEQ ID NO:757); GLDYVSHEDILPYTST (SEQ ID NO:758); DVHRETTENIRVTVIPFYM (SEQ ID NO:759); WWRYCIRLENLSDSDVQLRER (SEQ ID NO:760); and/or PAFQYSS
- 20 HVSLQASSGHMWGTFRFR (SEQ ID NO:761). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in gall bladder, prostate, and fetal brain, and to a lesser extent in a few tumor and fetal tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth related disorders such as cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 30 of the above tissues or cells, particularly of the prostate, gall bladder, and fetal brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

[REDACTED]

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth-related disorders, such cancers.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 184**

In specific embodiments, polypeptides of the invention comprise the sequence:SLCCPEGAEGC (SEQ ID NO:762) and/or QLKKTTHYDRPCP (SEQ ID NO:763). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in stromal cell, tonsil, and glioblastoma and to a lesser extent in some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
15 not limited to, immune and inflammatory disorders and glioblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, tonsil, and glioblastoma expression of this gene at significantly higher or lower levels
20 may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, it is believed that the
25 product of this gene regulates pancreatic cell differentiation into beta cells. Accordingly, polynucleotides and polypeptides of the invention are useful in the treatment of insulin-dependent diabetes mellitus and associated conditions e.g. pancreatic hypofunction and the prevention, as well as the treatment of undifferentiated type pancreatic cancers. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 417 as
30 residues: Pro-27 to Ala-32.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune and inflammatory disorders and glioblastoma.

35 **FEATURES OF PROTEIN ENCODED BY GENE NO: 185**

This gene is expressed primarily in hepatocellular carcinoma and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 418 as residues: Gly-32 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in hippocampus and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

less extent in osteoclastoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone-related disorders and neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, osteoclast, and hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone related disorders and neuronal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 188

This gene maps to chromosome 4 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 4.

This gene is expressed primarily in neuronal tissues such as hippocampus, spinal cord, and hypothalamus and to a lesser extent in a few other tissues such as ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 189

This gene maps to chromosome 10, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 10.

5 This gene is expressed primarily in neuronal tissues and immune tissues, and to a lesser extent in a few other tissues such as skin tumor, lung etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal and immune-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal and immune-related tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 422 as residues: Pro-19 to Asp-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal and immune-related disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 190

The translation product of this gene shares sequence homology with human N33, a gene located in a homozygously deleted region of human metastatic prostate cancer which is thought to be important in prevention of prostate cancer. In specific embodiments, polypeptides of the invention comprise the sequence:

30 AQRKKEMVLSEKVSQLEWTKRPMVIRMGDKFRRLVKAPPRNYSVIVMFTA
LQLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFAMVDFDEGSDVFQMLNM
NSAPTFINFPKAGKPKRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNMA
ARWRFWCVSVT (SEQ ID NO:765); MVVALLIVCDVPSAS (SEQ ID NO:766);
AQRKKEMVLSEKVSQLEWTKRPMVIRMGDKFRRLVKAPPRNYSVIVMFTA (SEQ ID NO:767);

MAVLSKMTLVV (SEQ ID NO:768); MVVDFDEGSDVFQMLNMNSAPTFINFPKAGKPKRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNMA (SEQ ID NO:769); KRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNMA (SEQ ID NO:770);

(SEQ ID NO:771); and/or YAGPLMLGLLLAVIGGLVYLRRVIWNFSLIKLDGILLQLCVLCLL (SEQ ID NO:772). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in infant adrenal gland prostate cell line and to a lesser extent in a few other tissues like liver, smooth muscle etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate cancer and endocrine disorders. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and adrenal gland, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 423 as residues: Pro-34 to Gly-43, Arg-113 to Pro-120.

20 The tissue distribution and homology to N33 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for prostate cancer and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 191

25 This gene is expressed primarily in T cell and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
30 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
35 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 424 as residues: Trp-3 to Phe-9.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 192

This gene maps to chromosome 6, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 6. Neural activity and neurotrophins induce synaptic remodeling in part by altering gene expression. This
10 gene is believed to be a glycosylphosphatidylinositol-anchored protein encoded by a hippocampal gene and to possess neural activity. This molecule is believed to be expressed in postmitotic-differentiating neurons of the developing nervous system and neuronal structures associated with plasticity in the adult. Message of this gene is
15 believed to be induced by neuronal activity and by the activity-regulated neurotrophins BDNF and NT-3. The product of this gene is believed to stimulate neurite outgrowth and arborization in primary embryonic hippocampal and cortical cultures and to act as a downstream effector of activity-induced neurite outgrowth. In specific embodiments, polypeptides of the invention comprise the sequence: DAVFKGFSDCLLKLGD (SEQ
20 ID NO:773); CQEGAKDMWDKLRKESKNLN (SEQ ID NO:774); VLLVSLSAALATWLSF (SEQ ID NO:775); MGLKLNGRYISLILAVQIAYLVQAVR AAGKCDVFKGFSDCLLKLGD (SEQ ID NO:776); PAAWDDKTNIKTVCTYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLNIQGSLFELCGSGNGAAGSL LPAPVLLVSLSAALATWLSF (SEQ ID NO:777); and/or MGLKLNGRYISLILA VQIAYLVQAVRAAGKCDVFKGFSDCLLKLGDXXXXXPAAWDDKTNIKTVCT
25 TYWEDFHSCTVTALTDCQEGAKDMWDKLRKESKNLNIQGSLFELCGSGNGAA GSLLPAPVLLVSLSAALATWLSF (SEQ ID NO:778). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in human placenta, endometrial tumor and tissues of the central nervous system (CNS).

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, relating to reproductive disorders, cancers and neurological diseases. Similarly, polypeptides and antibodies directed to the polypeptide of the invention are useful

in the diagnosis and treatment of the above disorders. In particular, in reproductive and neurological disorders, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 425 as residues: Asp-47 to Asp-63, His-75 to Tyr-80, Pro-83 to Tyr-89.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive disorders such as endometrial tumors. Expression of this gene in tissues of the CNS and its strong homology to Neuritin suggest that the protein product from this gene may also be used in the treatment and diagnosis of neurological disorders and in the regeneration of neural tissues, e.g., following injury.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 193**

The translation product of this gene shares sequence homology with tenascin which is thought to be important in development. The translation product of this gene is believed to be a ligand of the fibroblast growth factor family. FGF ligand activity is known in the art and can be assayed by methods known in the art and disclosed elsewhere herein.

This gene is expressed primarily in endometrial tumors, and other types of tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 426 as residues: Gly-29 to Glu-34, Arg-71 to Arg-76, Thr-176 to Cys-182, Gly-184 to Glu-199.

The tissue distribution and homology to tenascin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 194**

In specific embodiments, polypeptides of the invention comprise the sequence: MNSAAGFSHLDRRERVCLKLGESFEKQPRCASTLC (SEQ ID NO:779).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal human lung and neutrophils.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lung development and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
20 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal lung and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
25 and treatment of lung and immunity related diseases, for example, lung cancer, viral, fungal or bacterial infections (e.g. lesions caused by tuberculosis), inflammation (e.g. pneumonia), metabolic lesions etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 195

30 This gene is expressed primarily in breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunal disorders. Similarly, polypeptides and antibodies directed to these

tissues or cells, particularly of the immune system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunal disorders.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 196**

This gene maps to chromosome 5 and accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 5. The translation product of this gene shares sequence homology with human M-phase phosphoprotein 4 which is thought to be important in phosphorylation and signal transduction processes. In specific embodiments, polypeptides of the invention comprise the sequence:

15 TIYPTEEELQAVQKIVSITERALKLVSD (SEQ ID NO:780); RALKGVLRV
 GVLAKGLLLGRDRNVNLVLLC (SEQ ID NO:781); ALAALRHAKWFQARAN
 GLQSCVHILRLDLCQRVPTWS (SEQ ID NO:782); GDALRRVFECISSGIIL (SEQ
 ID NO:783); LAFRQIHKVLGMDPLP (SEQ ID NO:784); and/or TIYPTEEELQAVQ
 20 KIVSITERALKLVSDSLSEHEKNKNKEGDDKKEGGKDRALKGVLRVGVLAKE
 LLLRGDRNVNLVLLCSEKPSKTLLSRIAENLPKQLAVISPEKYDIKCAVSEAAIIL
 NSCVEPKMQVTITLTSPHIREENMREGDVTSGMVKDPPDVLDRQKCLDALAALR
 HAKWFQARANGLOSCVHILRLDLCQRVPTWSDFPSWAMELLVEKAISSASSP
 QSPGDALRRVFECISSGIILKGSPGLLDPCCKDPFDTLATMTDQQREDITSSAQFA
 25 LRLLAFRQIHKVLGMDPLPQMSQRFNIHNNRKRRRDSDGVDGFEEAGKKDKK
 DYDNF (SEQ ID NO:785). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Hippocampus and to a lesser extent in Prostate, Human Frontal Cortex.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders related to reproductive system and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
 35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and nervous system, expression of this gene at significantly higher or lower

levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human M-phase phosphoprotein 4 indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and nervous system disorders.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 197**

In specific embodiments, polypeptides of the invention comprise the sequence:

MGSQHSAAARPSSCRKQEDDRDG (SEQ ID NO:786);

ILAEREQEEAIAQFPYVEFTGRDSITCLTC (SEQ ID NO:787); and/or

QGTGYIPTEQVNELVALIPHSDQRLRPQRTKQYV (SEQ ID NO:788).

15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Primary Breast Cancer and to a lesser extent in Human Adult Spleen, Hodgkin's Lymphoma I, Salivary Gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and immunal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 430 as residues: Ser-126 to Gly-138.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and immunal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 198

This gene is expressed primarily in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, blood cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of blood cell disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 199

This gene is expressed primarily in Human Ovary and Synovia and to a lesser extent in Human 8 Week Whole Embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 200

This gene maps to chromosome 8 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 8. The translation product
 5 of this gene shares limited sequence homology with collagen proline rich domain.

This gene is expressed primarily in CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
 10 not limited to, neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 433 as residues:
 20 Pro-35 to Asp-41.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological diseases.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 201

Translation product of this gene shares homology with a mammalian histone H1a protein. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: ARLNVGRESLKREMLKSQGVKVSSEPMGAR
 HSSWPEGAAFCCKKVQGAQMFPFRR (SEQ ID NO:789); ARLNVGRESLKR
 30 EML (SEQ ID NO:790); LKSQGVKVSSEPMGARHSSW (SEQ ID NO:791);
 AFCKKVQGAQMFPFRR (SEQ ID NO:792). An additional embodiment is the polynucleotide fragments encoding these polypeptide (See Accession No. pirlS24178) fragments.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in vital immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such

15 as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 202

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

25 tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for

35 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 203

This gene is expressed primarily in Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
 5 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, infectious disorders, immune disorders, and cancers. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 immunological probes for differential identification of the tissue(s) or cell type(s). For
 10 a number of disorders of the above tissues or cells, particularly of the immune system,
 expression of this gene at significantly higher or lower levels may be routinely detected
 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
 an individual having such a disorder, relative to the standard gene expression level, i.e.,
 15 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
 NO: 436 as residues: Thr-31 to Lys-36.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosis and treatment of infectious
 20 disorders, immune disorders, and cancers. Since the gene is expressed in cells of
 lymphoid origin, the natural gene product may be involved in immune functions.
 Therefore it may be also used as an agent for immunological disorders including
 arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as
 well as, antibodies directed against the protein may show utility as a tumor marker
 25 and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 204

This gene maps to chromosome 16 and therefore polynucleotides of the
 invention can be used in linkage analysis as markers for chromosome 16. The
 30 translation product of this gene shares sequence homology with lactate dehydrogenase
 which is thought to be important in lactate metabolism.

This gene is expressed primarily in human tonsils and to a lesser extent in
 Spleen, and Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, immune disorders, infectious disorders, and cancers. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune disorders, infectious disorders, and cancers, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 437 as residues: Gly-7 to Ser-12.

The tissue distribution and homology to lactate dehydrogenase gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, infectious disorders, and cancers.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 205**

The translation product of this gene shares sequence homology with Gcapi protein which is developmentally regulated in brain.

This gene is expressed primarily in placenta and endometrial tumor and to a lesser extent in several other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vasculogenesis/angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Gcapi protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorder or dysfunction of vascular system of tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 206

In specific embodiments, polypeptides of the invention comprise the sequence
 MPYAQWLAENDRFEEAQKAFHKAGRQREA (SEQ ID NO:799);
 VQVLEQLTNNAVAESRFNDAAYYYWMLSMQCLDIAQD (SEQ ID NO:794);
 5 PAQKDTMLGKFYHFQRLAELYHGYHAIHRHTEDP (SEQ ID NO: 795);
 FSVHRPETLFNISRFLHSLPKDTPSGISKVKILFT (SEQ ID NO:800);
 LAKQSKALGAYRLARHAYDKLRGLYIP (SEQ ID NO:796); ARFQKSIELG
 TLTIRAKPFHDSEELVPLCYRCSTNN (SEQ ID NO: 797); and/or PLLNNLGNVC
 INCRQPFIFSASSYDVLHLVEFYLEEGITDEEAISLIDLEVLRPKRDDRQLEICKQQ
 10 LPDSCG (SEQ ID NO:798). Polynucleotides encoding these polypeptides are also
 encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 15 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and
 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the reproductive and endocrine systems,
 20 expression of this gene at significantly higher or lower levels may be routinely detected
 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
 an individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 25 disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for treatment of male reproductive and endocrine
 disorders.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 207

This gene is expressed in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and
 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
5 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 440 as residues: Tyr-49 to Cys-54.

The tissue distribution indicates that polynucleotides and polypeptides
10 corresponding to this gene are useful for detection and treatment of disorders associated with developing lungs particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of lung tumors since the gene may be involved in the regulation of cell division,
15 particularly since it is expressed in fetal tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO.	NT Total Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT First AA of Signal Pep	AA SEQ ID NO.	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
1	HLHDS67	97979 03/27/97	Uni-ZAP XR	11	2526	427	2526	458	458	234	1	30	31
2	HLHDZ58	97979 03/27/97	Uni-ZAP XR	12	1131	1	1131	129	129	235	1	14	15
3	HLNMJ13	97979 03/27/97	Lambda ZAP II	13	941	39	941	62	62	236	1	44	45
3	HLNMJ13	97979 03/27/97	Lambda ZAP II	218	941	39	941	245	245	441	1	35	36
4	HLTEI25	97979 03/27/97	Uni-ZAP XR	14	843	1	843	155	155	237	1	19	20
5	HMSJX24	97979 03/27/97	Uni-ZAP XR	15	1018	1	1018	90	90	238	1	18	19
6	HNFEID65	97979 03/27/97	Uni-ZAP XR	16	661	1	661	76	76	239	1	28	29
7	HNHDX07	97979 03/27/97	Uni-ZAP XR	17	553	1	553	106	106	240	1	23	24
8	HNHGC82	97979 03/27/97	Uni-ZAP XR	18	869	1	869	101	101	241	1	21	22
9	HNHGO09	97979 03/27/97	Uni-ZAP XR	19	959	1	959	176	176	242	1	21	22
10	HOU/BE18	97979 03/27/97	Uni-ZAP XR	20	1446	1	1446	101	101	243	1	27	28
11	HOU/DL69	97979 03/27/97	Uni-ZAP XR	21	1471	579	1460	692	692	244	1	31	32

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HPM171	97979 03/27/97	Uni-ZAP XR	22	1402	242	1402	401	401	245	1	32	33	60
13	HPMGQ55	97979 03/27/97	Uni-ZAP XR	23	1047	1	1047	164	164	246	1	26	27	35
14	HPQAC69	97979 03/27/97	Lambda ZAP II	24	990	1	988	82	82	247	1	20	21	37
15	HPTBB03	97979 03/27/97	Uni-ZAP XR	25	1208	350	1173	398	398	248	1	29	30	210
16	HPTWA66	97979 03/27/97	pBluescript	26	1922	1381	1922	24	24	249	1	33	34	547
16	HPTWA66	97979 03/27/97	pBluescript	219	575	1	575	148	148	442	1	22	23	65
17	HPTWC08	97979 03/27/97	pBluescript	27	1951	1422	1874	219	219	250	1	19	20	299
18	HRGCZ46	97979 03/27/97	Uni-ZAP XR	28	3989	2635	3989		2748	251	1	16	17	39
19	HSAVU34	97979 03/27/97	Uni-ZAP XR	29	3735	2966	3735	272	272	252	1	30	31	594
19	HSAVU34	97979 03/27/97	Uni-ZAP XR	220	3018	1929	3018	26	26	443	1	1	2	156
20	HSDFW61	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	30	1667	59	1625	138	138	253	1	32	33	130
21	HSDGP60	97974 04/04/97	Uni-ZAP XR	31	1408	1	1408	285	285	254	1			20

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
		209080 05/29/97											
22	HSOA155	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	32	2031	1273	2031	1285	1285	255	1	29	30
23	HSQE084	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	33	971	13	971	91	91	256	1	19	20
23	HSQE084	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	221	968	8	968	86	86	444	1	20	21
24	HSXAM05	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	34	1792	369	1792	470	470	257	1	26	27
25	HSXAS67	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	35	896	1	896	96	96	258	1	32	33
26	HTDAF28	97974 04/04/97 209080 05/29/97	pSport1	36	912	1	912	38	38	259	1	22	23
27	HTEGQ64	97974	Uni-ZAP XR	37	1382	67	1382	271	271	260	1		

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209080 05/29/97												
28	HTGEC09	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	38	872	1	872	74	74	261	1	18	19	28
29	HTOAM21	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	39	812	1	812	41	41	262	1	30	31	43
30	HTPBW79	209511 12/03/97	Uni-ZAP XR	40	1515	118	1507	302	302	263	1	24	25	362
30	HTSEV09	97974 04/04/97 209080 05/29/97	pBluescript	222	1404	1	1265	92	92	445	1	19	20	415
31	HJPCD40	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	41	704	22	704		117	264	1	18	19	127
32	HTWBY48	97974 04/04/97 209080 05/29/97	pSport1	42	1094	1	1094	32	32	265	1	34	35	53
33	HTWC146	97974 04/04/97	pSport1	43	1821	892	1647	56	56	266	1	26	27	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
		209080 05/29/97											
34	HTXGI75	97974 04/04/97 209080	Uni-ZAP XR	44	1024	30	1024		167	267	1	20	21
		05/29/97											
35	HWTFBF59	97974 04/04/97 209080	Uni-ZAP XR	45	983	779	983	85	85	268	1	30	31
		05/29/97											
35	HWTFBF59	97974 04/04/97 209080	Uni-ZAP XR	223	707	488	707	514	514	446	1	41	42
		05/29/97											
36	HADAE74	97974 04/04/97 209080	pSport1	46	2421	664	1587	710	710	269	1		
		05/29/97											
37	HAGFB60	97974 04/04/97 209080	Uni-ZAP XR	47	840	1	840	97	97	270	1	30	31
		05/29/97											
38	HATFE60	97974 04/04/97 209080	Uni-ZAP XR	48	2432	1193	2246	1491	1491	271	1	17	18
		05/29/97											
39	HBMSN25	97974	Uni-ZAP XR	49	1742	1165	1742	1207	1207	272	1	23	24

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209080 05/29/97												
40	HCDAR68	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	50	1487	181	1455	325	325	273	1	35	36	56
41	HCEJ79	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	51	1328	251	1328	525	525	274	1			21
42	HMDAN54	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	52	1856	725	1853	928	928	275	1	33	34	50
43	HCECA49	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	53	1558	310	1408	393	393	276	1			1
44	HCEEC15	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	54	948	1	948	9	9	277	1	23	24	65
45	HCESE40	97974 04/04/97 209080 05/29/97	pBluescript	55	990	99	990	193	193	278	1	32	33	256

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
45	HCESEF40	97974 04/04/97 209080 05/29/97	pBluescript	224	1384	99	1384	193	193	447	1	32	33
46	HCFMV39	97974 04/04/97 209080 05/29/97	pSport1	56	1603	1	1296	96	96	279	1	29	30
47	HCMISX86	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	57	1052	5	786	12	12	280	1	28	29
48	HCNAP62	97975 04/04/97 209081 05/29/97	Lambda ZAP II	58	814	1	558	93	93	281	1	22	23
49	HCRAF32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	59	1215	257	1215		356	282	1	19	20
50	HCLDC07	97975 04/04/97 209081 05/29/97	ZAP Express	60	478	1	478	147	147	283	1	36	37
51	HCWBB42	97975 04/04/97 209081	ZAP Express	61	618	1	618	212	212	284	1	35	36

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
52	HDTAB05	97975 04/04/97 209081	pcMVSpout 2.0	62	751	1	751	257	257	285	1	21	22	32
53	HE2AV74	97975 04/04/97 209081	Uni-ZAP XR	63	780	283	780	433	286	1				16
54	HE2AY71	97975 04/04/97 209081	Uni-ZAP XR	64	588	21	588	169	287	1				16
55	HE2GS36	97975 04/04/97 209081	Uni-ZAP XR	65	774	272	774	445	288	1				37
56	HE2OF09	97975 04/04/97 209081	Uni-ZAP XR	66	1866	1313	1866	1596	289	1				11
57	HE6FU50	97975 04/04/97 209081	Uni-ZAP XR	67	1152	117	686	237	290	1	20	21		34
58	HE9HU17	97975 04/04/97	Uni-ZAP XR	68	2483	1577	2448	1620	291	1				14

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
		209081 05/29/97											
59	HE9ND48	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	69	536	1	536	83	83	292	1	36	37
60	HEBBW11	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	70	865	647	865		388	293	1	30	31
61	HELDY74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	71	932	1	932	201	201	294	1	17	18
62	HEMAE80	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	72	996	1	945	12	12	295	1	24	25
63	HFEBA88	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	73	785	464	785	356	356	296	1	29	30
64	HFGAB89	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	74	1069	196	1047	295	295	297	1	32	33
65	HFEVHY45	97975	pBluescript	75	831	1	831		89	298	1	30	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
66	HGBAJ93	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94
67	HGBBQ69	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	77	1274	1	1273	105	105	300	1	24	25	43
68	HHFCF08	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	78	1133	4	1042	175	175	301	1	23	24	30
69	HHFHJ59	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	79	661	1	661	192	192	302	1	29	30	112
70	HHFHR32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	80	1378	1	1378		358	303	1			13
71	HHGCN69	97975 04/04/97 209081 05/29/97	Lambda ZAP II	81	1440	298	1440	532	532	304	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
72	HHGDOI3	97975 04/04/97 209081 05/29/97	Lambda ZAP II	82	1381	766	1371	993	993	305	1	23	24
73	HHPF63	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	83	1706	182	1644	257	257	306	1	24	25
74	HHSEG23	97976 04/04/97	Uni-ZAP XR	84	573	1	573	160	160	307	1	18	19
75	HIPAV06	97976 04/04/97	Uni-ZAP XR	85	684	199	684	323	323	308	1	27	28
76	HKIXL73	97976 04/04/97	pBluescript	86	1036	591	1036	690	690	309	1	32	33
77	HKMNC43	97976 04/04/97	pBluescript	87	908	1	908	139	139	310	1	18	19
78	HMEJE31	97976 04/04/97	Lambda ZAP II	88	655	1	655	165	165	311	1	33	34
79	HMSKS35	97976 04/04/97	Uni-ZAP XR	89	1102	1	1102	228	228	312	1	26	27
80	HINF4E54	97976 04/04/97	Uni-ZAP XR	90	1533	665	1518	347	347	313	1	26	27
81	HNFJH45	97976 04/04/97	Uni-ZAP XR	91	575	1	575	275	275	314	1	30	31
82	HNGBT31	97976 04/04/97	Uni-ZAP XR	92	639	1	639	224	224	315	1	28	29

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO:	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO:	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HNGIN60	97976 04/04/97	Uni-ZAP XR	93	744	1	744	225	225	316	1	43	44	70
84	HNGJG84	97976 04/04/97	Uni-ZAP XR	94	526	1	526	268	268	317	1	29	30	38
85	HNHDW42	97976 04/04/97	Uni-ZAP XR	95	426	1	426	168	168	318	1	28	29	71
86	HNHFL57	97976 04/04/97	Uni-ZAP XR	96	844	1	844	98	98	319	1	25	26	61
87	HOGAR52	97977 04/04/97 209082	pcMVSPORT 2.0	97	1985	453	1985	533	533	320	1	17	18	285
88	HOSBZ55	97977 04/04/97 209082	Uni-ZAP XR	98	1416	69	1416	246	246	321	1	32	33	54
89	HOSDI92	97977 04/04/97 209082	Uni-ZAP XR	99	1935	141	772		274	322	1	20	21	58
90	HPBCU51	97977 04/04/97 209082	pBluescript SK-	100	599	1	599	86	86	323	1	27	28	119
91	HPCAL49	97977 04/04/97 209082	Uni-ZAP XR	101	784	1	784		280	324	1	18	19	43

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
92	HPFCR13	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	102	1035	602	1035	859	859	325	1	32	33
93	HPHAC83	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	103	2218	840	2182	1035	1035	326	1	17	18
94	HPMBQ32	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	104	1351	1	1351	18	18	327	1	23	24
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	105	2066	51	2052	270	270	328	1	29	30
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	226	2057	1	1954	220	220	449	1	29	30
96	HRDFB85	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	106	1705	23	1697	233	233	329	1	21	22
97	HRGBR28	97977 04/04/97	Uni-ZAP XR	107	1167	611	1167	53	53	330	1	1	2

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	108	1907	151	1432	353	353	331	1	23	24	260
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	227	2084	335	2084	537	537	450	1	19	20	23
99	HSPAH56	97977 04/04/97 209082 05/29/97	pSport1	109	611	1	576	229	229	332	1	25	26	47
100	HE8EU04	209746 04/07/98	Uni-ZAP XR	110	2632	294	2632	337	337	333	1	25	26	333
100	HSXBT86	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	228	2143	53	1096	235	235	451	1			9
101	HSXCS62	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	111	2249	1	1953	90	90	334	1	18	19	199
102	HTEFL09	97977 04/04/97 209082	Uni-ZAP XR	112	2198	228	2158	400	400	335	1			23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
103	HTEKN135	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	113	1043	40	1043	320	320	336	1	20	21
104	HTGEP89	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	114	703	1	703	285	285	337	1	29	30
105	HTGEW91	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	115	3684	526	1338	584	584	338	1	24	25
106	HIOEY16	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	116	1965	127	1915	202	202	339	1	27	28
107	HTPCN79	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	117	503	1	503		1	340	1	7	8
108	HTSGM54	97977 04/04/97 209082 05/29/97	pBluescript	118	1133	316	1069		423	341	1	12	13
109	HTSHE40	97977 04/04/97	pBluescript	119	1101	118	956	218	218	342	1	31	32

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
110	HTWAF58	97977 04/04/97 209082 05/29/97	Lambda ZAP II	120	282	1	282	137	137	343	1	25	26	48
111	HTWBY29	97977 04/04/97 209082 05/29/97	pSport1	121	2635	1593	2489	1654	1654	344	1	25	26	55
112	HUKFC71	209007 04/28/97 209083 05/29/97	Lambda ZAP II	122	994	1	932	272	345	1	15	16	221	
113	HCE3Q10	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	123	1542	1	1542	143	143	346	1	25	26	63
114	HCEVR60	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	124	1390	82	1390	127	127	347	1	32	33	153
115	HDTAW95	209007 04/28/97 209083 05/29/97	pCMVSPORT 2.0	125	1288	412	1288	571	571	348	1			16
116	HE6EL90	209007	Uni-ZAP XR	126	1517	1	1452	243	243	349	1			9

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
		04/28/97 209083 05/29/97											
117	HELBUE29	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	127	1073	198	1073		776	350	1		
118	HERAH36	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	128	300	155	300	202	202	351	1		
119	HFXBW82	209007 04/28/97 209083 05/29/97	Lambda ZAP II	129	1275	1	1275	56	56	352	1	23	24
120	HHPD20	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	130	472	51	472		243	353	1		
121	HIBED17	209007 04/28/97 209083 05/29/97	Other	131	1950	284	1927	395	395	354	1	72	73
122	HLTER03	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	132	990	1	990	78	78	355	1	22	23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
123	HOABL56	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	133	1720	565	1720	660	660	356	1	18	19	21
124	HPMCJ92	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	134	705	28	705	106	106	357	1	28	29	98
125	HPWAZ95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	135	323	1	323	88	88	358	1	27	28	78
126	HRGBR18	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	136	582	1	582		16	359	1	17	18	30
127	HSTBW09	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	137	1021	1	1021	153	153	360	1	32	33	56
128	HUKC064	209007 04/28/97 209083 05/29/97	Lambda ZAP II	138	1777	439	1777		521	361	1			2
129	H6FAA53	209007 04/28/97 209083	Uni-ZAP XR	139	643	303	643		313	362	1	7	8	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
130	HAGAI11	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	140	1220	1	1220		127	363	1	16	17
131	HAGAO39	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	141	721	1	721		415	364	1		
132	HALSK07	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	142	1468	125	1468	210	210	365	1	29	30
133	HALSQ59	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	143	300	4	300	101	101	366	1	22	23
134	HAIBP89	unknown 05/18/98	Uni-ZAP XR	144	2243	173	2243	311	311	367	1	27	28
134	HBGCB91	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	229	1025	409	1025	624	624	452	1	20	21
135	HBMTD81	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	145	1082	163	1082	357	357	368	1		

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
136	HBXGK12	209008 04/28/97 209084 05/29/97	ZAP Express	146	4313	1153	4313	1313	1313	369	1	18	19	42
137	HFKFJ07	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	147	1183	1	1183	149	149	370	1	41	42	254
138	HCCQAI40	209008 04/28/97 209084 05/29/97	Lambda ZAP II	148	734	1	734	285	285	371	1			19
139	HCWHZ24	209008 04/28/97 209084 05/29/97	ZAP Express	149	1405	1	1405	108	108	372	1	34	35	63
140	HE2GT20	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	150	2890	1178	2890	1178	1178	373	1	31	32	39
141	HE8EY43	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	151	2399	1181	2399	1265	1265	374	1	30	31	34
142	HFCEB37	209008 04/28/97 209084	Uni-ZAP XR	152	802	352	802		487	375	1			10

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
143	HFTCT67	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	153	461	24	461	145	145	376	1	37	38
144	HGLAM46	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	154	2388	818	2388	648	648	377	1		
145	HHGBR15	209008 04/28/97 209084 05/29/97	Lambda ZAP II	155	642	322	642	400	400	378	1		
146	HJAAU36	209008 04/28/97 209084 05/29/97	pBluescript SK-	156	1251	583	1251		933	379	1	16	17
147	HUSIT49	209008 04/28/97 209084 05/29/97	pSport I	157	2127	247	2127	383	383	380	1	47	48
148	HKLAB16	209008 04/28/97 209084 05/29/97	Lambda ZAP II	158	1625	817	1625	1012	1012	381	1	18	19
149	HLMNU76	209008 04/28/97	Lambda ZAP II	159	1687	1307	1687	1296	1296	382	1	28	29

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209084 05/29/97												
150	HMSKQ35	209008 04/28/97 209084	Uni-ZAP XR	160	1842	172	1463	319	319	383	1	30	31	33
		05/29/97												
151	HNHED86	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	161	770	1	770	30	30	384	1	31	32	46
152	HNHEJ88	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	162	519	1	519	242	242	385	1	17	18	24
153	HNHHQ63	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	163	753	1	753	164	164	386	1	17	18	67
154	HOECT183	209009 04/28/97	Uni-ZAP XR	164	1400	189	1400		508	387	1	22	23	33
155	HPTRC15	209009 04/28/97	pBluescript	165	2153	594	2153		611	388	1			13
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	166	1251	219	1120			389	1			
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	230	1250	223	1250	393	393	453	1	32	33	171
157	H6EAE26	209009	Uni-ZAP XR	167	882	48	882	155	155	390	1	33	34	153

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
158	HAGBX03	209009 04/28/97	Uni-ZAP XR	168	1208	1	1208	182	182	391	1		
159	HAGDQ47	209009 04/28/97	Uni-ZAP XR	169	1307	1	1307	44	44	392	1	22	23
160	HAICP19	209009 04/28/97	Uni-ZAP XR	170	1624	89	1483	128	128	393	1	18	19
161	HAUAE83	209009 04/28/97	Uni-ZAP XR	171	2003	889	2003	1080	1080	394	1		
162	HBHAD12	209009 04/28/97	Uni-ZAP XR	172	786	1	786		176	395	1	17	18
163	HBNTY28	209009 04/28/97	Uni-ZAP XR	173	1758	962	1758	1184	1184	396	1	27	28
164	HBNTV04	209009 04/28/97	Uni-ZAP XR	174	888	330	862		546	397	1		
165	HCDDB78	209009 04/28/97	Uni-ZAP XR	175	2379	750	2379	901	901	398	1	18	19
166	HCEQA68	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	176	1348	1	1348	12	12	399	1	28	29
167	HCEZS40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	177	1502	178	1502	315	315	400	1		
168	HCFNF11	209010	pSport1	178	1637	26	1607	152	152	401	1	44	45

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	179	2911	1103	2858	192	192	402	1	32	33	424
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	231	1811	20	1811	93	93	454	1	36	37	95
170	HCUBL62	209010 04/28/97 209085 05/29/97	ZAP Express	180	519	1	519	57	57	403	1	28	29	32
171	HDSAP81	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	181	968	320	968	476	476	404	1	27	28	79
172	HE2CT29	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	182	1128	1	1128	111	111	405	1	26	27	94
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	183	2276	48	2276	88	88	406	1	37	38	257

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	232	2271	56	2232	79	79	455	1	43	44
174	HE9FB42	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	184	2500	76	1693	518	518	407	1	1	2
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	185	1337	60	1328	175	175	408	1	39	40
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	233	1338	33	1327	175	175	456	1	32	33
176	HEMVC19	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	186	941	33	931	79	79	409	1	23	24
177	HEMDX17	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	187	654	1	654	137	137	410	1		
178	HETAR54	209010 04/28/97 209085	Uni-ZAP XR	188	1848	454	1848	948	948	411	1	14	15

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
179	HETBX14	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	189	1146	157	1146	74	412	1	14	15	53	
180	HFGAB48	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	190	906	156	906	245	245	413	1	30	31	32
181	HEKFI40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	191	1941	120	1002	213	213	414	1	18	19	218
182	HFXHN68	209010 04/28/97 209085 05/29/97	Lambda ZAP II	192	2118	777	2118	966	966	415	1	23	24	50
183	HGBFO79	209011 04/28/97	Uni-ZAP XR	193	1538	259	1538	273	273	416	1	23	24	49
184	HGLAM56	209011 04/28/97	Uni-ZAP XR	194	1098	68	1098	185	417	1	28	29	69	
185	HHLBA89	209011 04/28/97	pBluescript SK-	195	1001	1	1001	324	324	418	1	25	26	39
186	HHPDW05	209011 04/28/97	Uni-ZAP XR	196	1443	1	1443	246	246	419	1	21	22	21
187	HHPSD37	209011 04/28/97	pBluescript	197	1282	66	1282	171	171	420	1	19	20	37

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion
188	HHPSF70	209011 04/28/97	pBluescript	198	951	26	951		162	421	1	16	17
189	HHS.AK25	209011 04/28/97	Uni-ZAP XR	199	1740	1390	1740	1534	1534	422	1	19	20
190	HLASB53	209011 04/28/97	pBluescript	200	1707	401	1195	652	652	423	1	26	27
191	HJABZ65	209011 04/28/97	pBluescript SK-	201	779	1	779	23	23	424	1	26	27
192	HJPBB39	209011 04/28/97	Uni-ZAP XR	202	1617	188	1605	182	182	425	1	28	29
193	HLHSK94	209011 04/28/97	pBluescript	203	1974	1	1794	112	112	426	1	26	27
194	HLHTC70	209011 04/28/97	pBluescript	204	1057	229	1057	365	365	427	1	23	24
195	HLMTW92	209011 04/28/97	Lambda ZAP II	205	721	1	721	244	244	428	1	25	26
196	HLTCY93	209011 04/28/97	Uni-ZAP XR	206	2465	988	2465	1225	1225	429	1		
197	HLTDB65	209011 04/28/97	Uni-ZAP XR	207	1480	1	1480		371	430	1	15	16
198	HMSHM43	209011 04/28/97	Uni-ZAP XR	208	872	1	872	35	35	431	1	18	19
199	HMSHQ24	209011 04/28/97	Uni-ZAP XR	209	1779	16	1779	148	148	432	1	24	25
200	HNF.AH08	209011 04/28/97	Uni-ZAP XR	210	2110	592	2110	611	611	433	1	18	19

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
201	HNGAO10	209011 04/28/97	Uni-ZAP XR	211	938	1	938	107	107	434	1	27	28	30
202	HNGBE45	209011 04/28/97	Uni-ZAP XR	212	1551	1	1551	114	114	435	1	21	22	100
203	HNHAZ16	209011 04/28/97	Uni-ZAP XR	213	997	1	997	202	202	436	1	24	25	36
204	HNHCN159	209011 04/28/97	Uni-ZAP XR	214	1496	1	1132		165	437	1	28	29	41
205	HOSFN122	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	215	1308	501	1308		809	438	1			1
206	HPHAC88	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	216	1705	384	1705	549	549	439	1	23	24	24
207	HCDEO95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	217	999	608	999	273	273	440	1	22	23	54

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological complex, thereby facilitating the identification of the gene.

As a further alternative, antibodies which bind specifically to the protein encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

- 5 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources
- 10 using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

- 15 Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1
- 20 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra.*) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

- 25 In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results
- 30 shown in Table I.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO-Y which have an N-terminus that cleaves to yield a protein of the

present invention, the amino acid sequence of which is secreted protein (not output).

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

25 As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:

35 Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is at least 95% identical to the

sequence of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,

- 5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

- As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be
- 10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
- 15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
- 20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

- If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.
- 25 For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
- 30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
- 35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. Examples of such variants include:

(1993), reported variant KGF² proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

One embodiment of the present invention relates to a polypeptide fragment comprising an

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the

polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5 Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

10 Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final
15 preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

 Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins
20 facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG)
25 can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

 Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules
30 together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified,

 various human proteins such as hIL-5 can be fused to an Fc portion of the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

15 **Vectors, Host Cells, and Protein Production**

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the present invention may be

well known to those skilled in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

5

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

10

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

15

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

20

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

25

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

30

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

35

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix formation, originally described by Lee et al., Nucl. Acids Res. 6:3073 (1979), involves the formation of a triple helix between a polynucleotide and a DNA or RNA sequence. This triple helix structure is stable under physiological conditions and can be used to inhibit gene expression.

Polynucleotides of the present invention can be used as antisense inhibitors and effective in therapy.

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (^{125}I , ^{127}I), carbon (^{14}C), sulfur (^{35}S), tritium (^3H), indium (^{112}In), and technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

Antibody labels detectable by X-radiography include, for example, ^{112}In , $^{99\text{m}}\text{Tc}$, a radio opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human
5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments," (Chapter 13 in *Tumor Imaging: The*
10 *Radiochemical Detection of Cancer*, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene
15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to
20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired
25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such
30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a
35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat clotting disorders.

Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

5 A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, 10 differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, 15 glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune 20 inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

25 A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The 30 administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may 35 inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenström's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

- 5 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, 10 Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., 15 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, 20 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

- Similarly, bacterial or fungal agents that can cause disease or symptoms and that 25 can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, 30 Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Hemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, 35 and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

- Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

- Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

- A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g., osteoporosis, osteoarthritis, periodontal

bone loss, liver failure, pancreas failure, kidney failure, heart failure, etc.) (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase
5 regeneration of tissues difficult to heal. For example, increased tendon/ligament
regeneration would quicken recovery time after damage. A polynucleotide or
polypeptide of the present invention could also be used prophylactically in an effort to
avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel
syndrome, and other tendon or ligament defects. A further example of tissue
10 regeneration of non-healing wounds includes pressure ulcers, ulcers associated with
vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a
polynucleotide or polypeptide of the present invention to proliferate and differentiate
nerve cells. Diseases that could be treated using this method include central and
15 peripheral nervous system diseases, neuropathies, or mechanical and traumatic
disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and
stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral
neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized
neuropathies, and central nervous system diseases (e.g., Alzheimer's disease,
20 Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-
Drager syndrome), could all be treated using the polynucleotide or polypeptide of the
present invention.

Chemotaxis

25 A polynucleotide or polypeptide of the present invention may have chemotaxis
activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes,
fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial
cells) to a particular site in the body, such as inflammation, infection, or site of
hyperproliferation. The mobilized cells can then fight off and/or heal the particular
30 trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase
chemotactic activity of particular cells. These chemotactic molecules can then be used to
treat inflammation, infection, hyperproliferative disorders, or any immune system
disorder by increasing the number of cells targeted to a particular location in the body.
35 For example, chemotactic molecules can be used to treat wounds and other trauma to
tissues by attracting immune cells to the injured location. Chemotactic molecules of the
present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

5

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit
10 (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural
15 or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

20 Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing
25 the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results
30 in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring the amount of bound

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with
15 a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

20 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic
25 surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac
30 rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

35 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under
5 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which
10 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide
15 sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at
20 least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous
25 nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500
contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide
30 sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer
35 as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

- 5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of
10 comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

- A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

- 20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

- 25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

 The method for identifying the species, tissue or cell type of a biological sample comprises a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing

comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSPORT 2.0	pCMVSPORT 2.0
pCMVSPORT 3.0	pCMVSPORT 3.0
pCR [®] 2.1	pCR [®] 2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Altling-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Altling-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lacmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

The following is a list of the cDNA clones deposited with the ATCC:

Human CD44, GenBank Accession No. U01251, GenBank, Bethesda, MD (1992).

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then
5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA
10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic PI library (Genomic Systems, Inc.) is screened by PCR
20 using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by,
25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is
30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1100-1. Full length cDNA libraries are screened by the same

analytical procedures

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8. the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

The target fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

- 5 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

- 10 The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

- Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

- 20 The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

- 25 The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

- 30 Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

 To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Comassie blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The

express the cloned polynucleotide

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGold™ virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

- 5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)
- 10 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in
- 15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

- To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection
- 20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins
- 25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., *J. Biol. Chem.* 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., *Biochem. et Biophys. Acta*, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., *Biotechnology* 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five μ g of the expression plasmid pC6 is cotransfected with 0.5 μ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 μ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blotting.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

```
GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
```


AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
 AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
 ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
 GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
 5 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
 GAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCTGG
 ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA
 GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
 ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
 10 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of
 15 the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell
 25 Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at
 30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the

relatively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The
 5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x
 10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in
 15 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of
 20 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off
 25 PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1% BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl_2 (anhyd); 0.00130 mg/L
 30 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 0.050 mg/L of $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$; 0.417 mg/L of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 311.80 mg/L of Kcl; 28.64 mg/L of MgCl_2 ; 48.84 mg/L of MgSO_4 ; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO_3 ; 62.50 mg/L of $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$; 71.02 mg/L of Na_2HPO_4 ; .4320 mg/L of $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$; .002 mg/L of Arachidonic Acid; 1.022 mg/L of

Acid, 0.010 mg/L of Palmitic Acid, 0.010 mg/L of Palmitic Acid, 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0
 5 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-
 10 Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;
 15 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x
 20 penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B
 25 adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

30 It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an
 35 activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the

identification of a reporter molecule. The activation of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u> <u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>	<u>STATS</u>	<u>GAS(elements) or ISRE</u>
<u>IFN family</u>							
5	IFN- α /B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	Il-10	+	?	?	-	1,3	
<u>gp130 family</u>							
10	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	Il-11(Pleiotrohic)	?	+	?	?	1,3	
	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
<u>g-C family</u>							
20	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
	IL-15	?	+	?	+	5	GAS
25							
<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
<u>Growth hormone family</u>							
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
<u>Receptor Tyrosine Kinases</u>							
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCG
AAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATG
ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
CTAACTCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCCTAGGCITT
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively resulting in the SV40 promoter element linked to the GAS promoter element.

OTHER CELL SYSTEMS: The above described system can be used in a variety of other cell systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final
5 concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentecin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12
20 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples
25 from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

- 10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

- 20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

- 30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
 5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC
 TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTIGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC
 ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCCA
 TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCCATGGCTGACT
 AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes Sall and NotI, and inserted into a vector containing neomycin resistance. Particularly, the

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000-20,000 cells/well in a Co-star black

96-well plate. Wash the cells with Hank's Balanced Salt Solution (HBSS) (Gibco) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

- 5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100
10 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

- To measure the fluorescence of intracellular calcium, the FLIPR is set for the
15 following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

20

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

- The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase
25 RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

- 30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of

5 activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr

10 with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of

15 alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of

20 Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇

25 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum

30 manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many

determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 5 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 10 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- 15 Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as
- 20 above.

- Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
- 25 tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

- As a potential alternative and/or complement to the assay of protein tyrosine
- 30 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
- 35 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts
15 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C using buffer

polynucleotide kinase, employing SequinTherm Polymerase (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

5 The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

10 Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

15 Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

20 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

25 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If
30 given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

10 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

15 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

20 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

30 The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence in culture.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense
5 DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art,
10 see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290
15 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a
20 pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the
25 polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in
30 the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the
35 transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is

...antibiotic resistant DNA matrix ... the foreign DNA ...
either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 μ m cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

Sequence Listing

(i) GENERAL INFORMATION:

5 (i) APPLICANT: Human Genome Sciences, Inc., et al.

(ii) TITLE OF INVENTION: 207 Human Secreted Proteins

10 (iii) NUMBER OF SEQUENCES: 800

(iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

20 (C) CITY: Rockville

(D) STATE: Maryland

25 (E) COUNTRY: USA

(F) ZIP: 20850

30 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

35 (B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

40

(vi) CURRENT APPLICATION DATA:

45 (A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

50

(vii) PRIORITY APPLICATION DATA:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Kenley K. Hoover
(B) REGISTRATION NUMBER: 40,302
(C) REFERENCE/DOCKET NUMBER: P2007PCT

(vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG 60
AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCAAA ACCCAAGGAC ACCCTCATGA 120
TCTCCCGGAC TCTTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG 180
TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG 240
AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT 300
GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG 360
AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC 420
CATCCCGGSA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT 480
ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGAGAAC AACTACAAGA 540
CCACGCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG 600
ACAAGAGCAG GTGGCAGCAG GGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC 660
ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC 720
GACTCTAGAG GAT 733

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
 1 5

10

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 86 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CGGCCTCGAG AATTCCCCGA AATCTAGATT TCCCCGAAAT GATTCCCCG AAATGATTTC 60
 CCCGAAATAT CTGCCATCTC AATTAG 86

30

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTGTGCAAAG CCTAGGC 27

45

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 271 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

AAATATTTCG GATTCTAATT AATCAAAAG GATAATATTC TTTTAAATTA GAGATATTC 100

60

GCGCCTAACT CCGCCCACTT CCGCCCATTC TCCGCCCAT GGCTGACTAA TTTTTPAT 180
 TTATGCAGAG GCGAGGCG CCTCGCCTT TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240
 5 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10 (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

20 GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

25 (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 GCGAAGCTTC GCGACTCCCC CGATCGCCT C 31

40

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
 45 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

50 GGGGACTTTC CC 12

55

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs
 60 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 GCGGGCCTCGA GCGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60
CCATCTCAAT TAG 73

10

(2) INFORMATION FOR SEQ ID NO: 10:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(E) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTCTCCCGGG GACTTTTCGG GGAATTTTCG GSACTTTCCA TCTGCCATCT 60
25 CAATTAGTCA GCAACCATAG TCCCGGCTCT AACTCCGCCC ATCCCGCCOC TAACTCCGCC 120
CASTTCCGCC CATTCTCCGC CCGANGCCTG ACTAATTTTT TTTATTTATG CAGAGCCCGA 180
GGCCGCCCTCG CCGTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTS GAGGCTTAGS 240
30 CTTTTCCAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2526 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GACAGGCTAT CCGAGAATCT GAGAGCTGAG CCGGGCAATT CCTCCAGTTA CCGTTGTGAG 60
CTAAGTCCAG TCACACATTT CCCAAAGTTT CTCTTTGTCA TAACCCTGGT CTGGCTGGTT 120
50 TTGREGRCTT GACAATGGST CAGGGAATCC AGGCCAAGTC CAACAGAGAC CCCAAACCCA 180
CCACACACCA CGAGCCAGAA CCGAACAGCC AACAAAGAGG ACTTTTGTGG GAGCACAAGT 240

60 GAGGCTTAAT CAGCTTTTCT TCGCAATAT TATATATATG ATTATATATG GAGAGAGAT 480

	GGCCAGGCCC CCAGCGACTC TTCTTGCCCT GATSTTTGTC CTCACAGGCA TCCCACCTGG	540
	CCTGAGATGA TTCAGAACAA ATCATGCTAA CTCTGAATCC ATCCAGCCAC TTGCAAATGA	600
5	TAATCAGAAG TCAGCTTGTT CACTGTTAGA AAGAACTAA CAAAAGAGAA CCCAGAGCAA	660
	TCTAGAATCT TTGAGTGCTT GGCTTTCCAA GGATACTGCG GAGATCTCGG CCAAGCTCAT	720
10	GACCTTCTGA ARTGTCACTG GCACCATATG CAACAAGAAC CACCAATCAC TGAGTAGCTA	780
	ATGGGTCTCG GGCTGGGAC ATTCCATCTG AGGTCCCTCC TGAACATGTC ACTCCACAGC	840
	AGAGGACCGG TTGCAGCTTA CCCAGAACCA CTCCTCCAGG AGAGCTGAT GTTTTGGCTG	900
15	CAACACCTTG AGCACTCACT GCTATTGTTT AAAAAAGCC TTGTGTCAT TCGGAGGACT	960
	GGCCGCTGCC CTGAGGTGAC TTCCTAATA TGTGGTTTCA TTAGCGAAT TATTTTCTG	1020
20	GCTGGCTGGA CATTTGTATT TTGTTAGGTT GCTGTTTAAG CTCAGTTTG CTGTCTCTC	1080
	TGCAGCTACA AAACATCTTG CATATTAA GAKTGGCTTT TATAAATAGC TTTATTCTGA	1140
	TATTAATGAG ATTCCCAACT TCACTGAGAA TTAAGGACTG GGTACTTTA AAGAAATGCA	1200
25	AATAGCAATT GAAGAACCAC TCTTCAGGT GSTAGCCCTG CCTAGACTGA ATTACATAG	1260
	AAATCAGCCA GAAGGAAGCG TCCTTGGGAT CCCAGATCAC TCTTTTTTTT TTTTTTTTA	1320
30	AAAGGGGCG CCGCTTGATG GCTCATCTCT CTGAATAACA GTTACGTCTT CATATCGATA	1380
	CCAGATGCTT TCTTCATCAT GCCACTGAAG CCACTCACCA CTTTCAAGAA CATGCCAACC	1440
35	TCTGTGAGAT TCACTTACCC ACAAACAAGG AGGCACGTTT GGCACAAAGT GTTGTCTCTT	1500
	AGSTCCAAGT GGAATCTACA GAGTGCTTGA CCTCAACACA CTGATTCACA GGTGCACTGG	1560
	ACCAAGAGCA GCCAAGACA CCGGAAGTGA AAAACTCCAC AGGCTTTGSA GAATAGAAAT	1620
40	GAAAAGCCAC GTCATATAAC TCAAGAATAA ATGGTGTTTT GGAAATTTTA AAATTATCAT	1680
	CGAAGGTGCT GAAACTATTT CAGGCCCAAA TGAAAGGAAA TCGCCAGTTG GGGATGAAAT	1740
	CACAGAGCTT GTCTTTTATG ATATGGTTGG ATGTCCACTG ATGAAATTTT AAAGGAGTTT	1800
45	CATTTTAAAA AGTGGCGATG ATTCTACATA TGAGAATTCT TTAGGCCAAG AAAGTGTCTT	1860
	TGGCTCAGAG GTGTTTGGAA TTAATGCAGA GAGAAGCCAT TCGTGATGCT TAGAACCAAG	1920
50	GATGCTCATG TACACAAAGA CCATCGAGAC GGCCATTCTT GTTTACAAAA CACTTACCAA	1980
	GAAAGCACTT TGTAGGGGAA CTTTAGTAAG TTCTTCTCAT TTCAATTATG TTCTTCCAAG	2040
	GAAACAGGAG AGACTGAATT AATAATTCTC TCTTTCCTCT TAAGCACTTT TAAAATAATA	2100
55	AAGTACATCT TGAATTTGG GGGGTCATCT CTGATTTAAA AAAAGAAAAA GGCTGCTTGA	2160
	TGTATGTTAT GCAGAGACAC TCTGCTCTG GTGGCTGCAG AGCAATACCC AAGCCTCAT	2220
60	TGGAAGGCTC AACATTTGGA ATTGCACCTT AATGATTAA TCCTCAATTC ATGTGCCCTT	2280

ACGGGATGGT GGGTCTGGGA CCCCAATTCA TTCTTATCTG CCAAAGAATT ATCTAGAAGC 2340
 ACATCAANTA CCAGCACCCC ACCTGCACAA TGGGGGTGGA AAACTTTGT ATCCCTAAGC 2400
 5 ATATTATTTT ATAGTGTCTG CCATGCCATG TGGAAATACT TTATTTTAA CTTCAGGATT 2460
 TAAATAANGT AAACACTATG ACATTTAAAA AAAAAAAAAA AAAACTCGAG GGGGGCCCGG 2520
 10 TACCCA 2526

15 (2) INFORMATION FOR SEQ ID NO: 12:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1131 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(3) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 CACTGCACCA GCTTTGTTAT CTGTAAAATG ATGATAATAC CAACACCTTC TTCTTGGGGT 60
 ACTGAAGATG AGAGAACATG ATATGTGTAA AGTGCCTTCC ACAATACCCA GAACATAGCA 120
 AACATSTAAT GAATGTAGTA ATAGTAATTA TTTTATTTTC TTTTGATTCA GTTGGGACTA 180
 30 TGTTCAGCTG TAACAGAATA CCCAAAATAA CTGTTTAAAT CAAATTAAGT TTTWSTTGTG 240
 AAGTTTGTG ACGAATTCAG ACAATCCAGG GCTTTTATAG ATGCACCAGG ATCAGCAGGT 300
 35 ACAAAGGAT CTTTCTTGAT TTCTGCCAGT CTCAATGCAT GGGTTGCAAT CCAGARTCCA 360
 REATGGCACT TCCAGCCCTG GTTAGGCCCA TATTAGCACA CAGAAAGAAA GAGAAAGGGA 420
 TGTGCTCTTT CACTTTAATC ATAGCTCCCA CTAGATGCAC CCACTACTTC TGCTGATACT 480
 40 CCATTAGTA ATGCTTGCTT ACATGGTCAC ACTTAGTTTC CAGAGAGACA TGTCTGGACA 540
 GTCATGCTCT CAATTAATAT CCAAGTGTCC AATTACTGAG AAAAAAGAA ACTAGCAGCT 600
 45 TTGCTTGCTT GCATTCCTCT TAGCATAAGC CACATTCTTT TTATGAAGTT GTCTCAGTT 660
 AATTGATGC CTGAGTGTTC CTTTCAWITA GAAAGGCTTC TGGGACATCT TAAAGCTGAC 720
 TTCTTTTGTG ATCAGCACCA TCACTACCAC TGCCCTCTTC AAAGCCACCA CGTTCTGTCC 780
 50 CCAGGATGGT TGCAACAACC ACCATAGGGA CTTTTTCCT TCTACTTCCA CACAATAGNC 840
 CAGAGTAAGC TTTTGAAGT GTAGGTGAGA TCATGTCTCT CTCTTCTCT TCAAAAGCTT 900

60 THE SEQUENCE OF THE INVENTION IS IDENTICAL TO THE SEQUENCE OF THE INVENTION 1.0

TATTTCGACT TAAATAGAA AAAAAAAAAA AAAAAAACT CGAGGGGGGC C

1131

5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 941 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

15

GGCAGGAGTA GCAATTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCATGT 60

GATGTCTCTA GCGAAGATTG TTAGGCAGAG AGGAGCTCTC CCAACCTACT ATACCACCGA 120

20

GGCTGGAGAG ATTATATTTT TGSTATATAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA 130

TGTCCTCTGT AGCAAAACCGG AAAAGTCAGT GACAGAAGAT GCGGCTAGCG GTTTGAGCCA 240

GAGAATGACA GCTCTGTTTT GGAGAAAAGG GCGGATGCTT GCTCTAGAA AGCCCATCCT 300

25

TCTGCTCTTC TTTTCTCTCC CCTTATATT GTGCTTTCAT TCATTCATTC ATTCATCAAA 360

CATTTGTTGA GCACTATTA TGTCTCAAGC TCTGTCTAG CCTCTGAAA ACCTGCCCTC 420

30

ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGGTTGTCA 480

GGGTCTCACA GAGCACTGGC CCTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC 540

GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC 600

35

TGAAGGTGGC ACTGCTTGA GTCTTGATTC CAGCAGAGGG AGAGCACTCT GTGAAAAGGC 660

ACCAAGGGTG GGAGAGGCA GACCACATGG AGGAACCTCA GGTAGTTCTG GATGGCCTG 720

40

GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA 780

TECCAATAAA CCTATTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTAA GGCTGGTCCG 840

AGTAGAATGA TTTTACAAC GAATTGATCA CAACCAGTTA CAGATGTCTT TGTTCCTTCT 900

45

CCACTCCCAC TCTTCACCT GACTAGCCTT TAAAAAAAAA A 941

50

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

55

(A) LENGTH: 843 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

60

CNAGGGATAA CCCCAAAGNT GGGAAATAAA CCTCAATTA AAGGGGGAAC CAAAAAGCTG 60
 GGAAGTTCCC GGGGGCGGTG GGGGCCNGNT CTAGGAACTA GTGGAATCCC GGGGGGCTCC 120
 5 AGGGAATTGG GCAAGGAGTG GGAATGTTGT TTGTATGATA CTATTTCAC AAWATGCATT 180
 GAGACTTGGT KTGTGGCTA GGACATGCTC AATTCTTTT AAATATTGCG TCAATTCTT 240
 TAGTGCATAT TCTCGATGG GGGTGTGGG GACAGAGTTC TAAATATGCC CATTAGATTA 300
 10 AATCTCTTCA TTCTGTGTCT CACATCTTCT ATATCTTAT TAATCTGTCA ATCTCTTCAA 360
 GAGAGGTGTT ATTAAAATCT CTCACTGTAT GTGTCACTTT GCCCTFAAAA TTCTGATGAT 420
 15 TTGCTTTATA AATGTTATA ACCATTTTCC AGGAAGAACA TTAAAGAACT TTCCATTGGC 480
 ATTATCCAST TTCCCTEAAA ATACTGTMTT TTTTATTTT GCTTCTAAG CAGCTATGAA 540
 TCCAGTTTCT CAGAAGCCTT TGTCTCAAGG CATTTGTTTC CAGATTACCT TGTTAGCATC 600
 20 CACACTATGG GCTATTTTAG AAAAACAAAA AAAGTATCAA AATCATATAG CTATGATTTT 660
 CCTGTGCTTG AAGGAGCTT AAAGCTCATC TAGTCCAGCC AGTATTTGTT CATCCAAATT 720
 25 CTGCCAAGAA ATCTCTATG TCAAGATATT CTTTACCATC TTTGGGACAT TCTCATTATT 780
 AGAAACAAAT CCTAAGAAGA AATTCTGCCA TAKACAACCC ATCCGTTCTT TAAAAAATAA 840
 AAA 843

35 (2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1018 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45 CTGTAATTTT TAAATTTGAT ATACCGTCTT TGGATTCTAA TTTTATTTT TGAGTTCTCT 60
 GAAGGTTACA TATACAGACT GCTTCAAGAA TGATCATTTT GTTATTATTC ATATCTCTTA 120
 ACAATGTTCT TTTACTGCAA GAAGATAATT GCTAGAGAAA GAATACAGTG CAGGAAAGAA 180
 50 GARGCTGGAG CCACTGGTGA AGARGGATTG AGARGACAGA CATGTGCGGA ATGAAATCAT 240
 GAATAATGCT GTTTTGAAT TGTCCAAAAA CTTCTACAAA CCATGAAATG TTGAGTTTAA 300
 55 TATGATTTT TATGATTTT TATGATTTT TATGATTTT TATGATTTT TATGATTTT 360
 60 TATGATTTT TATGATTTT TATGATTTT TATGATTTT TATGATTTT TATGATTTT 420

274

5 TGGCTATCTA ATTTGSTGCC AAATACTTAA TGTGCTTGAA TTAAAAACA GCAAACATGT 600
 AGAAAGGTAA TTATAATAT GAGGCTAGTT CTTTAAGCTA GCTTTTTTTC CCTCTCAAA 660
 CAGCATATTS GTTGGATST CAGCAGGAGA AAGTSTTTTT TCCAATACAC ATAATGCATA 720
 TATGCTCTCTG TTAGCAATCT ATAGAAAATA GATATTGCTC ATTAAGGTAA ATATTTTGT 780
 10 TGATGAATGA TGTGGAATGS TGTGACTTG TTGTGTGAAC AGGAAATTGC TGTGTAGGCT 840
 TTGACTTGTG AGTAAAGAG TGAGGTGCT AAGATTAATT AAAGTAAATA CTGTGACAAT 900
 ACGATGTGAA AACTAAAAAC GGTGTTCTGA AACTCAAGGA ATTAATGACA CATAGGGAAG 960
 15 TTTTTCCTAT ATTAAGCATA GAGTAGGAGA CGCAAGTCAA GAATAAAAAA AAAAAAAA 1018

20

(1) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 661 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

30

35

40

45

50

55

TTTAAGAAAT TACTGAATTC CCGGTCAG GGAATTCGGC ACGAGGAGGA GGCCGTCAGC 60
 TGGCAGGAGC GGAAGATGAC AGCTGTTGCC CCGGGTTGCA CCCCCCAGY TGTGCTGGAC 120
 ATAAGTGTCT TACAGAGAGC CCGGAGGT GGGGAGGCTG TACCTGTGGA GTGCCGGCAC 180
 CGCCTGCAGC TGGCTGAGGC AAGGAAGGGG CCTCTGAGCC CAGCATGGAT GCCTGCCTAT 240
 GCCTGCCAGC GCGCTAGGTC CTTACACAC CACAACACTG GCCTMTCCGA GCTGCTGGAG 300
 CATGAGTGT GTGAGGAGST GGAGAGAGTT CCGCGCTCAG AGAGGTACCA GACCATGAAG 360
 GTGCCCAGGG CAGGGCTGG ACCTACCCCA GGAATGTCCT GCCCTGGGAA TGACAACACA 420
 GTCCACACCA TGTACGGGGA GGCAAACAGG GGCAGCTGAC CCAGCCCAGG GGTGAGANGA 480
 GGTCTTGCCG AAGAAAGTGC AGCTAAGCTG ATACCTGATA TGCACWAGKC AGCCARGYGG 540
 AGACAGGCAA GGAAGAAGCT TGTGTTGAGG ACAGAAATTT CTAGATCACT CAGCACCATC 600
 TGGCTTTTGG GGTGTTTTGT TTTATTTTGT TTTTGAGAGG GGGTCTCGCT CTGTGCCCCA 660
 N 661

(2) INFORMATION FOR SEQ ID NO: 17:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 553 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ IL NO: 17:

	GGCAGAGGGG TATTTGCCCC TCTCTCCACA TGACAGAACT GCTCTAAGTT TCTTTGCTGC	60
10	TCTTCTCAGC TGTCAGAGGG CTTCCTGCTT GTTTTCCACA CCACCATGTC TATTCTTTGC	120
	TCTCTTTCAC TCTGCTGTT TTTTCTCTTT TGTATTTCCT CTGCTCTCTG TCCCTTTTCC	180
15	CACGTGTGWC AGCTTTCTTT TATTGCCACT TTEAGTCAGA GCAGTCTGT GTTCTCTGGT	240
	CCGECATACA ATACTTACTT GAGTTTCTTG GCTTTTCTTG ACTGTGATC TCTTACTTCA	300
	ACATAGGAAT AGCTGTGAT AGAATTTCCT CACTTCCAGG GCTCAAGAGG GAGAGTCCCA	360
20	GAAAAATTGAG ACTSTTTTCC CTGCTTTGGA TTGAATTCAT AAACCAAAAC CACTCTTTGT	420
	GTGAGGGCTTT GCTGTGTGAT GCTATAGGT TCTTTGCTG CAAACCTATA GAATCCAGCC	480
	TGCGAAAACA AAGRAACCAG AGAATANCAG CATCAGAACA ATGCTTGACA TCATTCTCA	540
25	ATCAAGCAAT CCA	553

30

(2) INFORMATION FOR SEQ ID NO: 18:

(2) SEQUENCE CHARACTERISTICS:

35

(A) LENGTH: 869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

40

40	GGCACCAGCCT GCCAACACTG AGGTCTTCGT GGCCTCTCAC ATCTAGATGT ATCCCTCTCA	60
	AATCTATCTT CTATCCAGCC ACACAGNTTGA GGTATCTAAA ATGTCAACTT TCCAGTTACT	120
45	CTTCTTATA CTACCCCAAT CAAGCTACAA GATAAAGTCC AAGCCCTTTC ATATGACAAA	180
	CCACACCTTG CTTAAGTCTC CAGGTTTGAA TCCCTCATCT CTTACTTTAA ACTTTAAAAAC	240
50	CCAGCAGCAC GAAAGTGTCT CTTATGCATG TTGCCATATG CTTTCTCTCC ATCATGCATT	300
	TGCTTGAGCA AGATGTCTTG AGTTAACATC TTATTCTTTA AACTTCATTG TGGTGGTAGA	360
	CAGCCTTTAA TAACGATCTC TTGGCTAGGC ACAGTGACTC AAGCTGTAA TCCGAGAACT	420

60

RECEIVED: 10/1/97

CAAGGCTGCA GTGAACCATG ATCAGAAAT TGCANTCAG CTTGGGTAAC AGACTGAGAC 660
 CTTAGGTEAG AAAAATGAAT AAATAAGAT AAAATTTTAA AACTTAGCC AGGCATGGTG 720
 5 GCACACATCT GTGCTCCCTG CTACTTAGGA GGCTGAGGTG AGAGGATCCT TGAGCCCAGG 780
 AGGTCAACAC TACAGTGAGC TATGATTGTG CCACTAAACT CCAACCTGG TGAAAAAGCA 840
 10 AAACCTTGGC AAAAAAAAAA AAAAAAAT 880

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 984 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GGCGAGCCGA GATCGTGCCA TTGCACTCCA GGCTGGCCAA CAAGAGTGAA ACTCTGTCTC 60
 25 AAAAAAAAAA AATTATAATA CTATATCCCA TAAATGACA TTTCATATTT AAAGAGTTTT 120
 TTAAAACTCT TGTATTACA TCCATAAT TGAAACCCTA TTTCAGTGAA TGAGAATGGT 180
 30 ATCTGTGTG CTCATTTTTT CATTTTATC CTTAACAATT TCCACCACAG CCAGTGCATA 240
 TAATGGCAAT GACACCCAGG GATGGAATGA TAAGTTCCAT CCGGCTCAG TCAAGACGCA 300
 GACTTGATGT GGGCCCAACA AAGTCAATA ATGGAGTCTC CAAAATAAAG CTCTATAGGA 360
 35 AAGGTAAATA CCGGTGCAC AAGAAACAC AGCATCTAGG TTCTAACCCC ATCTCTATGA 420
 AGAGCTTCTT GGGAGAGTTT TGACATTWAA CAATCTGTCT GATGCCAAT TTTTTCCTC 480
 40 TATAAAATGA TAATGTTKGA YTCAAAATC CAAAGTCAAT TCATGGTCTA AACTTAATG 540
 ATTTTTTTAG GTTTTGKGAC ATTTCACTGT AACTGTAGT AATTATATC TTATTTTCCC 600
 ACTAATTTAG AAAAATATYT AAATGATCCT TAATTGCCAA TGGGTCTTAA GAATTTTGT 660
 45 TTAAATCCCT GTTACCCAAA AGAGCCCTTT TTTGTATCTC GCAGTAGTTA CAAGGATCTT 720
 TCTAAATCTT AAAAAAAAAA AAAAAAATA GAAAGAAAAG AAAAGAAAAA AAGTCAGCCG 780
 50 GGGTGGTGG CTCATGCCTG TAATCCACG ACTTTGGGAC CAAGGTGGAC AGATCACAG 840
 GTCAGGAGAT GGAGACCATC CCGGCCAACA TGGAGAAACC CTGTCTCTAC TAAAAAAAAA 900
 55 AAAAACTCGA GGGGGGCCCG GTACCCAATN CCGCGCTAG TGGTCGTAAA ACAATCAA 959

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 1446 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CGGGGCAGGG CTGTGTGGCA CGGDEACGGA CGCGGCCAC CTGAGTCACT TTATTGGGTT 60
 10 CAGTCAACAC TTGTCTGCTC CCTGTTTTCT CTGTGTGGG ATGATCTCAG ATGCAGGGGC 120
 TGGTPTTGGG GTTTCTGTC TTGTGCAAG GCGTGACAC TGCTGGGGGG CTGGAAGGCG 180
 15 CCTGCTTGG TTTCTTCTG TGCGCTCAT GCGCTCATG GTGCTGCCAT CCTTCTGGA 240
 GAGAGGGAGG TGAAGCTG TGTAGCTTA GTGGGTTGCG GCGCACTCAC CCAGGAGTGG 300
 GCTGGGCCAG GAGGCGGAGA GCGAGCACTG CTGCGCTGCT GCGCGCTGCT CTTCGCGAGT 360
 20 TAGGGGTGGA CGAGGCTCG GTTCTGCGAC TGTGTGGAG CGAAGGGGAA GAGAGGGGTC 420
 TTCAGGCTCG AGCGAGCTG GCGGTGCTG GTGAGAGAT GAGATTTAGG GCGTGTCTCA 480
 25 TGGGGTGGC AGCGCTGGG TGAAATGGA AAGGCGCAGA ACGTGCAGGT CTGCGGAGGG 540
 GAAGTGTCT GAGTGAAGGA GCGGACTGCG ATCCTGGGG ATGCTGGGAG TGAGTGAAGT 600
 AGATGGCTGA GTGAGGTTA TGGGAGGCT GAGGTTTTAT GGGCTGTGT ATCCCTTCT 660
 30 CCGCGCGGCA GCGTGTCTG CTGCTGCTG CCTGGGCCAC AGGTGTGCT CTGCTGCTG 720
 TCCCTCTGT GTTTGGGAT GAGCGGCGAG CAAGGGGTGT AATGGGGTG GGTCTGTCT 780
 35 TTTACAGGC ACCCGAGGT CTCAGTGTG TGCTGCGGA GCGGAGGGG GCTCTGAGG 840
 GSTACAGTT GGTGAGGCG TCGTGAAGG TGTGGGTCA GCGTTTGGT CTGCTGCTG 900
 TCAGTCACCA ATCACTGCT CTCTGAAAT CCAGTGGCT GTTTGGATGT CCTGTGAGT 960
 40 CACTCTGGG CTGGTGTG TCCCTCTCA GTTCTTGT GTTGGGACAA GGTCAAGGC 1020
 AGGATGAGG GAGGCTGGG ATGCTGACG CCAGGAGGCG GAGGCGCTT CCGCTGCTG 1080
 45 TTTGGGAGG GAGGAGGCA AATGAGTCT TTTTGGGTC GCGAGGTGG GTTCTGCTG 1140
 AGCGCTGAT GTTCTGCTG GAGGCTGCT TGTGAGAG AGGGGCTTG ACCGATGGA 1200
 CGTGCTGCTG GCTCTGCTG TAGGAGGCG CCAGCTGCTG GAGGCTGCT CTGCTGCTG 1260
 50 TCTCTGCTG GTTCTGCTG AGATGCTG CTCTAGTCT TTTTGAAGG TTCCATCAT 1320
 CCGTCTGCT TATTGATTG AAAATATTAT GCACTGCTT CAGCTTCTA CTAATCAATA 1380

60

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1471 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAAAAAATAA TAATGATAAT TTAAAAATAA TAAGTAACTA ATAAAAAGAT TTTATATCCC 50
 AGTCTTATGA TGTGCTGTTGG CAAGGCTAGA TAAAAAGATG TTAGAATGAA AGAACATATT 120
 15 TTTACTGATA TGTAATGAA GGATTCTACA ATAGTCATAT ATTTTATAT GAATGAATGT 180
 TGGGTGCGG TCGAGAGTA TGTGTGTGTA AATATAAAG TCTCACATTC AGAGTATAGC 240
 20 TCTGAAATAA TGGAACTCAT GTGTACAATT CAACATGAT CTGTATAGTT ACATCTCATG 300
 TAAATATACA CAGACATATT TTGCAAGCAG TAATTGAGAG TTAATGTCCA AAACAGGTGA 360
 TTGATAGGTA ACAGAAATTA GATAACCACC AATTTTCCCC AAGAGAAAGA CTAGAAGGAC 420
 25 TAAAAGCACT TGAATGTATG CTACTGACAT TGTGATAAGC AGTCTGATAA CCAGTTTATT 480
 GAAACGTGTG CATTAACAGA GAATTTAATT TAAACCCAT AATTTCTCCT ATCCATTAAA 540
 30 ATATTATAAT TGTTAGTAT ATGAAACCAA CAGGAAATCT TTTTAAATCA TTTAGTGAGG 600
 TGATTATTT GTTTCATGG CAAACACTAT CCAAGAAAAG CCTTGCTTGC CTGTTTCCCA 660
 AAGAGTCTA AGAAATAGAA TCAAGTGTA AATGGTTCAG ACCATTGAGG ATTTCTGTG 720
 35 ACTCTTCTCA ACCCGATCT TCCTGTATT ACTGATGTT GAAACCTGT CATTAGCCCC 780
 GGCCTGTTA AAGCCCTCA GAGTCACCTC TCATTCATAG CAATAGAATT CAACCCCAAG 840
 40 TGGTTSATGG TGTCCCGAGC ACAGCCGAGA GACCTGATCT CTGGATTGAG TGCTTTTAGC 900
 TCTTCGAGTT TACCCTAAGA TACCTTCGGG CAATATTTTT AACCAACCCA AAAGCTCTTC 960
 ACGTCATTTG TGAAGAGGAC AAGGTGAATC TTGGCTTGGA ACACCATTTT TGGGCTCTTG 1020
 45 CTACTGAATG AATCAGAAAG GAATTTTTC TGAAGAGCAT TAGAAAGTAA AGGAGATGTT 1080
 AAAATAAGTT CTGAAAGTAT GTTTTATATT TATCTAAAAC ACTGATTTTA AAASTTTACA 1140
 50 TTCAAATGTG TATTCAAAAG AAGTACTGAT TTGTAATTAT TATAGTTTGT GTGTATCATC 1200
 CCCTTTTAAC CGTGCCTAAC AACTGTACTT AAATTTTGTT TTCCTAGTGT AACAAATGTT 1260
 TCCATAAGA TTTTCTAGAG CCAAATAATG GGAGTGAAAA ATTCTTAAG TGTATATAA 1320
 55 GAAATATAT TAGAAATCA GCTTTGGATT ATACGATTTT TAAATATAC TAATACAGAA 1380
 TCCTCAGTAA TATGTTTTGA ATTGGATTTT TTCTCAGAAC TGTTACATAA TAAATAATAC 1440
 60 ATCAACCAGA AAAAAAAAAA AAAAAAATTN C 1471

ATAAAAAAAAAA AAAAAAAAAA CT

1400

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1047 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

15

GGCACAGGGG ACTACAGGGA CCCACGACCA TACCCAGCTA ATTTTGTAT TTTTGTAG 60

AGATGGGGTT TCACGATGTC GCCCAGGCTG GTCTTGAACCT CCTGGGCTTG AGCGATCTTC 120

20

CCATCTTTCC ATCTTGGGCT CTAAAGTCC TGGGACTGCA GGCATGAGCC ACCATGCCCA 180

GCCAAGATTC TTATTCATTA CATGTTGCT TCAAGAAGCC AAGCCAGTTT CCAATATTC 240

25

CCATTGCTG GAGTCTTGT ACTTTGGSTA GAAGCAACTG GTAAATTGTT AATTGGAACA 300

NTTGGTGGTG TATATAACCA CGTATGGGCA AACCTAGAGC ATCTAGGCTC ACAATTACTA 360

TCCTGACTTG ATAACAAGTG TTCTGATATT AACCTGAAAA TGGAATAAT GCCAAATCTG 420

30

TGTAACCTAA CATCTATATA CACAGTGGGG AGAACTGAAG TTATTAAACC TCGAATCTCT 480

GTGATCAAGG CTAACAGTAG TTATCTAAGA AGCAAAGGAC CTACAATTCT TAGACTTGGA 540

35

GTCATATTC TTAAGGACCT GTTCTGAAAC TATATCAAGC ATCTGGTTTC CACGTATTTT 600

TCCCTCAGAA ATTATGAAGT ACAAGTAAAA ATGAAGGTAC ACGGTAAGAC ACATGCTGCT 660

TTCTTGCTCT TGAGTGAGA CAGTTTCCCA GCCATCTTAA CCCCTTWACA CAAAACAATT 720

40

TGTGTTTTAT AGCAAATAAG TGAATCAACA TAATTTCAAT ATGATGTTTA TCCACCAGTA 780

CTTTCCTTTC AGCTTCTAGT CCCATAARTG GTTGTGAAG TCATCGGTTA CATTAGCCAA 840

45

GATAGGCTTA GACTTGAAGT CTAGAATGTT TTTCCCACTA TATGCCAAAG TAGAATGTGG 900

GTATCTCAGG GTCATTTTTG TTGTTCAATT TCCCACCTGT ACAGTTGTTA TGATTCACCT 960

TCCTTATGTG TCTAATAAAT CTGTTCCAT GAAATGATCA AAAAAAAAAA AAAAAAACT 1020

50

CGAGGGGGGG CCCGGTACCC AAATCGC 1047

55

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(2) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 TTGGAAAGGG TCTAGCTCTT TCTCATTCAC CAACTATATT AGAAGCACTT GAGGGAAATT 60
 TACCACTCCA AATCCAAAGC AATGAACAGT CTTTCTGGA TGATTTTATT GCCTGTGTCC 120
 CAGGATCAAG TGGTGGAAAG CTTGCAAGGT GGCTTCAGCC AGATTCATAT GCGGATCCTC 180
 10 AGAAAACATC TTTGATCCTG GAATAAGGAT GATATTCGTT GTGGTTGGCC TACCACCATA 240
 ACTTTCAAA CAAAAGACCA GTATGGGLAT GTGTACATG TTCCCAATAT GAAAGTAATT 300
 15 ATAACTGGAT TAAATTAGTA GACATCTATA TACTGCTGC AATGACTGAT AAAATTTTAG 360
 AAATGCCAAG TGCTGAGGCT CCATTTGTTT TACCTCTTT ATATAAAGCG TGATGCTGAA 420
 AGTTTGTFTA AATGACTTGT TTATATTAAT TATCCCCAA GTGTCCAAGT TACACCTGTT 480
 20 TTTTGTGTA GTTGTGTTCT TACATTTTGC TACCTGTTAC GGGGACTCAA AGGAGCGATA 540
 AGAAAGTATC CATCTAAAGA GTCTAGACA CATACAGTGA AGCCCTCAA TATGTATTGA 600
 25 TTGAATAAAT GCATGAAAGA ATACATTTT AAATTTGTG TATAGTTTGT AAAGACTCAA 660
 GTACGTTCTG TTTTTGGTAT TACTGAAACC ACATTTTAAA AATAACACTC ATTAAGTTAG 720
 AAATATATGA GTTTAGATTG TAAAGAAATG AGGAATTGAA ATAGTTGTAT ACCATATTGA 780
 30 TGAATATAGA GTTTTATAGA TACCTCTTAC CTGAAATATT AATAATAATG TTTTCAGAGC 840
 ATATTATACA TAATTATTG TGATTTAATC TGTTAATATG AATATCTCAT TTAACCTTT 900
 35 TATTCTGAA AAAATTATAT TGAATAAAAT TTTATATAGG CAGTCCCCAG CCCTTTCCTC 960
 CTTCAAAGTT CTCTTATAGA GTGATTGTT 990

40

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1108 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TAATCGCTAC TATAGGGAAA GCTGCTGCT GCAGGTACG GTCCGGAATT CCGGTCGAC 60
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 120
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 180
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 240
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 300
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 360
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 420
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 480
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 540
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 600
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 660
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 720
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 780
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 840
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 900
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 960
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 1020
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 1080
 CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC CCGGTCGAC 1108

60

CTCTTGGGCT CCTGAGCTCC AGGCGGTGCG CATGTTTGCT GACTAAGTCG CCCACGAGAG 360
 TCGAGGGGAC AGCATGCTGG CCGAGCTGGA CCGAGAGATG AGCAGGAGCK TGGACGTGAC 420
 5 CAACACCCACC TTCTGCTCA TGCGCGCTC CATCTATCTC CAGGACAGA ACCCGGATGC 480
 CGCCCTGCGT GCGCTGACCC AGGCGGACAG CCTGAGTGC ACAGCCATCA CATTGAGAT 540
 10 CCTGCTGAAG CTGGACGGC TGGACCTGG CCGGAAGGAG CTGAAGAGAA TGCAGGACCT 600
 GGAAGAGGAT GCCACCTCA CCCAGCTGC CACTGCTGG GTCAGCTGG CCACGGTGG 660
 TGAGAGCTG CAGGATGCTT ACTACATCTT CCAGGAGATG GCTGACAACT GCTGGCCAC 720
 15 CTTGCTGCTG CTCAATGGC AGCGGGCTG CCACATGGC CAGGGCGCT GGGAGGCGG 780
 TGAGGGCTG CTGAGGAGG CCTAGACAA GATAGTGGT TACCCGAGA CCTGGTCAA 840
 20 CCTCATGCTC CTGTCCAGC ACCTKGGCAA GCGCCCTGAG TGACAAACG GATACCTGTC 900
 CCACCTGAAG GATGCCACA GGTCCCATC CTTCATCAA GATACAGG CCAAGGGAG 960
 CGACTTTGAC AGGCTGGTC TACAGTACG TCCAGCGCT GAGGCTGGC CAGAGCTGTC 1020
 25 AGGACCATGA AGTCAGGACA GAGGCCAGG GCGAGCCTG CAGCCCTCC CACCCGGCAT 1080
 CCACCTGCAT CCTCTGGG CACGAGCCA CCGCCAGCA CCGCATCTGT TAATAATAT 1140
 30 CTCAACTCCA GGTGTGCA CCTGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200
 AAAAAAAAAA 1208

35

(2) INFORMATION FOR SEQ ID NO: 26:

40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1922 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTGCTGCGCT ACTGAGCAGC GGCATGGAGG ACTTGAAGC ACTGGGCTTC GAACACATGG 60
 GCGTGGATCC CCGCTCCTT CAGGCTGTCA CGATCTGGG CTGGTGGGA CTTACGCTGA 120
 50 TCCAGGAGAA GGCATCCA CTGGCCCTAG AAGGGAAGGA CCTCTGCTT GGGGCGGCA 180
 CGGGCTCCGG GAAGACGGC GCTTATGCTA TTCCGATGCT GCAGCTGTG CTCCATAGGA 240
 55 AGGCGACAGG TCCGTTGTA GAACAGGCAG TGAGAGGCTT TGTTTGTGTT CTTACCAAGG 300
 AGCTGGCAAG GCAAGACAG TCCATGATTC AGCAGCTGGC TACCTACTGT GCTGGGATG 360
 TCCGAGTGGC CAATGTCTCA GCTGTGAAG ACTCAGTCTC TCAGAGAGCT GTGCTGATGG 420
 60

	AGAAGCCAGA TGTGGTAGTA GGGACCCCAT CTGCGATATT AAGCCACTTG CAGCAAGACA	480
	GCTTGAAACT TCGTCACTCC CTGGAGCTTT TGGTGGTGGA CGAAGCTGAC CTCTTTTTTT	540
5	CCTTGGCTT TGAAGAAGAG CTCAAGASTC TCTCTGTCA CTGCCCCG G ATTACCAAG	600
	CTTTCTCAT GTCAGTACT TTTAACGAGG AGTACAAGC ACTCAAGGAG CTGATATTAC	660
10	ATAACCGGT TACCTTTAAG TTACAGGAGT CCGAGCTGTC TGGGACAGAC CAGTTACAGC	720
	AGTTTCAGT GGTCTGTGAG ACTGAGGAAG ACAAATTCCT CTTGTGTAT GCCCTGCTCA	780
	AGCTGTATT GATTGGGGGC AAGTCTGTGC TCTTTGTGTA CACTGTAGAA CGGAGTTAC	840
15	GGCTACGCT GTTCTTGGA CAGTTACGCA TCCCTACCTG TGTGTCTAAT GGAGAGCTTC	900
	CACTGCGCTC CAGGTSCCAG ATCATCTCAC AGTTCAAGCA AGGCTTCTAC GACTGTGTCA	960
20	TAGCAACTGA TGCTGAAGTC CTGGGGGCCC CAGTCAAGAG CAAGGTGCGG GGGCGAGGG	1020
	CMAAGTGA CAAGGCTCT GATCGGAAG CAGTGTGCG CCGGGGATA GACTTCGAC	1080
	ATGTGTGTGC TGTGCTCAAC TTGATCTTC CCCCACCCC TGAGGCTTAC ATCATCGAG	1140
25	CTGGCAGGAC AGCAGCGGCT AACAAAGCAG GCATAGTCTT AACCTTTGTG CTCCGACGG	1200
	AGCACTTCCA CTTAGGCAAG ATTGAGGAGC TTCTCAGTGG AGAGAACAGG GGGCCCATTC	1260
30	TGCTCCCTTA CCAGTTCGCG ATGAGGAGAG TGAAGGCTT CCGTATCTC TGAAGGATG	1320
	CCATCGCTC AGTGAATAAG CAGGCCATTC GGGAGGCAAG ATTGAAGGAG ATCAAGGAAG	1380
	AGCTTCTTCA TTCTGAGAAG CTTAAGACAT ACTTTGAAGA CAACCTTAGG GAGCTTCAGC	1440
35	TGCTCGGGA TGACCTACCT TTGCACCCC CAGTGTGAA GGGCCACCTG GGGCATGTT	1500
	CTGACTACCT GCTTCTCTCT GCTCTCGTG GGTGTGTG CCGTCACAAG AAGCGGAAGA	1560
40	AGTGTCTTTC CTCTGTAGG AAGCCCAAGA GAGCAAAGTC CCAGAACCCA CTGCGTAGCT	1620
	TCAAGCACAA AGGAAAGAAA TTCAGACCCA CAGCCAAGCC CTCTGAGGT TGTGGGCTT	1680
	CTCTGAGCT GAGGACATTG TGGAGGACAG ACTTACCTC TTCTGTGACA GCGGAGCTC	1740
45	TGTTCTTAC TGTACAGCT GAGCAGACAG TTCTGGGCG GGTATCTG GGGCTTTAG	1800
	CTCTTGGCA CTCTAAGCT GGCATCTTGC GCTTTGACAA CAGATAAAA ATTTTACTG	1860
50	CCCCAAAAA AAAAAAAAAA AAAAAAATC GAGGGGGG CCGTACCCAA TTGCGCTAT	1920
	AA	1922

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5	TGGTCCCTAAG AGGTGGCTGA GCGCCAGGAG SAGGGTGAGG GGGGAGCCTG GGGAGCCCG	60
	CGCCACCTCC AGCGGCTCT CTGAGCTGG ACACCAGGGC CTTCTCTAT GATCTGTCA	120
10	AGTACAGCT GGTGTAGAT GAGCATGCAC AGTGGAGCT GTGAGCCTG CAGCTGCTT	180
	CGTAGACTAC AGTAGGAGA GTGATCTCT CAGCTCTAT GACAACTGT CTTGCTCTC	240
	CTGCGCTTAT GATCTGCTA TGGTAAGGA ATATGAGGAG GTGCTGCGG CAGAGCGCC	300
15	TGCTGCTCT TCGAGGAAAC TCCAGCTCT ATGAACCGA GGTCTATTTC TCGAGAAAT	360
	TCTGAACGT YTTGATGAT GCGGCTCTC GTCTCTGAG TCTGAGTCT TCGAGCTGT	420
20	TCTCTGAT CATCAAGGG GAGGAGGAG AGCAGACCTA CCGGCGATA TTAGGATTG	480
	TGCTCGACA CGAAGAGAA CTTGAGCTG AAGTGAAGA CTTCTCTCA GTGAGCTCC	540
	AGCTGAAGA CTACTGTAC GAGGCTTACA ACATCGGAC TGGTGGCGG GCTTCTTTTC	600
25	CTGCTTATTA CCGATTCAG CTCACCAAG AGCGGAGCA CATGCGAGC CTGCTAAAA	660
	ACAGTCACTG GTTGAACAG TCCGGCTCA ATTCTCTGG CTCATTCAG CTTCTTATC	720
30	ACAAGCGCAA TGACTCTCT TGTCTGCTA TCGAAAAGAT TGGCAGGAC CCGGCTCTA	780
	CCCTGCTCT TAAAGCGCC TCCAGCTCT TCTGGAGAT CAGCTGCGG GGTCTAAGA	840
	TAGGCTCAA GCGGATGAG TCCAGGAGG CCAAGGCGA TAAATGTAG CACTTTTTC	900
35	AGTTAAAAA CATCTCTTC TCGGATATC ATCAAAGAA CAGCAAGTAC TTTCTCTCA	960
	TCACGAGCA CCGGCGGAC CAGCGCTTG CTTGCCAGT CTTGTGTCT GAGACTTCA	1020
40	CCAAGCCCT GGTAGAGTC GTGGGAGAG CATTCCAGCA GTTCTACAAG CAGTTTGTG	1080
	AGTACACTG CTTACAGAA GATATCTAC TGGAGTAGT GTGAGCCCC GCTCTCTGC	1140
	TCTCTAGTC CTAGGCGAG TCCAGGAGA GTTGGCTGT GATAGGATGT GGTCTGCTT	1200
45	GAGGAGGGG ACTTGTGAC GCGAGAGAG AAGGAAGTG GCGCTTGGC CAGGTAGGG	1260
	GAGGCTGGG CAATGCGAG AGGCAATGC AGTTTATTGT AATATATGG ATTAGATTCA	1320
50	TCTATGAGG GCAGAGTGG CTGCTGTGG ATTGGGAGG ACAGGCTTG GGGAGCAGT	1380
	CTCTGGCAGA GAAGGATTC CTTTCAAGG GCACAGGCG CTGCGCATC CTGGGCTTA	1440
	CTTCCCTGC CAGGCTTGG GCGTGTGGC TCTTGCCTG ATGAAGCGG TCTCTGCTT	1500
55	TGATGAAGC TGTGCCACT GCAAGTGGC GCTTGGGCG TGCCCAACC CCGACGAG	1560
	AGCCCTGAG TCAGGCTGAG CCGAGCAGC TCGAAGGAC TTTTCAGTGA GAAATGGCA	1620
60	ACAGCTGGG GTGAAGTCC TGTCTCAGC TCGTCTATC GCGGGCTTC TGGTGGCTC	1680

285

CTGCGCACTGA CCTCAGCGGC ATGCTGGCTT GTGGCAGGCC TAGGACCTCA GCGGGGGAAK: 1740
 AGGAGCTGCC GCAAGGCTCT GTGCGAGCAG AAGAGGGAGG CTTCCTGACT GACACAGGCC 1800
 5 AGGCGCATCT TGGTCTCTTC AGCTTGCTTC CAACTATTAA AGTCCCATTT CCTGTCAAAA 1860
 AAAAAAAAAA AAAATGGGGG GGGGTCGGGA ANCCAATTTC CCGCAAAAAG GGGGGTTATA 1920
 10 AAAATTCGCG GCGNGCTTTT TIAAAAATTC G 1951

15 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3989 base pairs

(E) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25 GGCACAGGCC GCAGGGAACC TATGGGGGCA TATAGCTTGT AATGAAACTG TAGTCTCAGT 60
 TGGAAAGCTA GACATGAAAT GGTCTAGTGA GCAGGGCTCT ATTCTTAGTC TCCAGGCAATG 120
 CCTGTGGAAC CTGAGGCTTC TCTAGCACA TTGAGGCGAG GCAGATGYAA AAAATTCACA 180
 30 GAACTATGAT TGGACTTCAA GGTCTTGTAG ATTCTCTCTT TCATTCTAAT TTCAGTGTCT 240
 AAAATCTTTG CATCTTCAA CAGCTGGGCC ATTTGATGAG AGAGGGCTGA ATACTGCACT 300
 35 TTCTCTCTTA GAAATCATCT GGGGCATTTT CTTCGAAGTG AAGGGAACAA TAAGGCATAA 360
 CTCTTTTCAC AAATTTGGGA TAATGATTTT TGGGATAAAG ATCTACGAGA ATAGGGATAT 420
 TTCAGCTTTG GTTTGAGAT GAAAGGCAAA GAATATCATG ACCAGCTTTC AGGCTCTCTG 480
 40 AAGTATATCT CTCACATTGT CTTGTTCTCA TGCTGAGGAG CTTGAGATTC CTGTGTGGGG 540
 ATTAGATAGT GGATCTTAT GGTGTAGCTT GAATGCTTTT ATTTCTCTG TCTCTCTCTG 600
 45 AATGTATCTT AGGAATCAA AAGGAGGAG AAAGGAGGA AGAGGAGGCT CAGCAATCTT 660
 CAGGCTTGA GAGGAGCTG CTGAGGTAG TAGAGCTTGA AGTCTTCTAG GACTACTCTG 720
 ATAGATGTTA TCAACTCTT TCAATTTCTT TGAACAGCC TGAATCTCTC CAGGCTTATG 780
 50 GAATCTCTTT TTATGCATTG GAGGAAAAAC ATCTTGGCTT TTCTCTTGAC GTGGGAGAAA 840
 TTGAAAAGAA GGGGAAGGGG AAGAAAAAAG GGGGAAGAAG ATCAAAAAGG GAAAGAGAG 900

60 ATTGATATAT TCGAATCTT TCAATCTT TCAATCTT TCAATCTT TCAATCTT TCAATCTT 960

	GSAGCAACAG CATCTTGGGT TGGCTGTGTA CATGGATGAA ATPGAAAAGT ACCAAGAAAT	1140
	GGAAGAAGAC CAAAGACAT CATGCCCCAG GCTCAGCAGG GAGCTGCTGG ATGAGAAAGA	1200
5	GCTGAAATC TTGAGGACT CACTGGATAG ATGTTATTCG ACTGCTTCAG GTTATCTGA	1260
	ACTGCTGAC TTAAGCAGC CTTACAGCAG TGCKGTTTA TCATTGGAGG AMCAKTACTT	1320
	TGGCTTHKCT CTTCACGTG ASAAATTGAA AAGAAGGGGA AGGGGAATAA AADAAGGGGA	1380
10	AGAAGATCAA AGAAGSAAAG AAGAAGGGGA AGAAAGSAG GGAAGSAGS TCAAAACCGA	1440
	CCATGCCCCA GGCCTAGCAG GGAGCTGCTG GATGAGAAAG GGCTGAAAT CTGCGAGGAT	1500
15	TCACTGGATA GATGTTATTC AACTCCTTCA GGTGTCTTTG AACTGACTSA CTCATGDCAG	1560
	CCCTACAGAA GTGCTTTTA YRTATTGGAG CAACAGYGTG TTGGCTTGGC TGTGACATG	1620
	CATGAATTC AAAASTACCA AGAAGTSSAA GAAGATCAAG ACCCATCAAG CCCAGGCTT	1680
20	AGAGGGGAGT TCCTGATGA GAAAGAGCTT GAAGTCTTGG AGGACTCAAT GGATAGATGT	1740
	TATTCGACTC CTTCAGGTTA TCTTGAAGTG CCTGACTTAG GCGAGCCTA CAGGAGTGCT	1800
25	CTTTACTCAT TCGAGGAACA GTACCTTGGC TTGGCTCTTG ACCTGGACAG AATTAAAAAG	1860
	GACCAAGAAAG ACCAAGAAGA CCAAGGCCCA CCATGCCCCA GGCTCAGCAG GAGCTGCTG	1920
	GAGGTAGTAG AGCTGGAAT CTGCAAGSAC TCACTGGATA GATGTTATTC AACTCCTTCC	1980
30	AGTTGTCTTG AACAGGCTGA CTCTGDCAG CCCTATGGAA GTTCCTTTTA TGCATTGGAG	2040
	GAAAAACATG TTGCTTTTC TCTTGAGTG GAGAAATTG AAAAGAAGGG GAAGGGGAG	2100
35	AAAAGAAGGG GAAGAAGATC AAGAAGRAA AGAAGAAGGG GAAGAAAGA AGGGAAGAA	2160
	GATCAAAACC CAGCATGCCC CAGGCTCAAC GCGGTGCTGA TCGAAGTGA AGAGCTSAA	2220
	GTCTTACAGG ACTTACTGGA TAGATGTTAT TCGACTCCGT CAATGTACTT TGACTACCT	2280
40	GACTCATTCG AGGACTACAG AAGTGTGTTT TACTCATTTG AGGAACAGCA CATCAGCTTC	2340
	GCCCTTTACG TGGACAATAG GTTTTTTACT TTGACGGTGA CAAGTCTCCA CTGGTGTTTC	2400
45	CAGATGGGAG TCATATTCCC ACAATAAGCA GCCCTTASTA AKCGAGAGA TGTCACTCCT	2460
	GCAGGCAGGA CCTATAGGA MGTGAAGATT TGAATGAAAG TACAGTTCCA TTTGGAAGGC	2520
	CAGACATAGG ATGGGTCACT GGGCATGGCT CTATTCCTAT TGTCAAACA TGCCAGTGGC	2580
50	AACCTGTGCT CAGTCTGAAG ACAATGGACC CACGTTAGST GTGACAGST CACATAACTG	2640
	TGCAGCACAT GCGGGAGTG ATCAGTCRGA CATTTTAATT TGAACCAST ATCTCTGGST	2700
55	AGCTACAAAA TTCTCAGGG ATTTCATTTT GCAGGCATGT CTCTGAGTTT CTATACCTGC	2760
	TCAAGGTCAK TGTCATCTTT GTGTTTAGCT CATCAAAAG TGTACCCCTG GTTCAATGA	2820
60	ACCTAACCTC ATTCMTGTG TCTTCAGTGT TGGCTTGTTT TAGCTGATCC ATCTSTAACA	2880

CAGGAGGGAT CCTGGGCTGA GGATTGTATT TCAGAACCAC CAACTGCTCT TGACAAITST 2940
 TAACCCCGTA GRCCTCTTG GTTAGAGAAG CCACAGTCCT TCAGCTCCA ATTGGTGCA 3000
 5 GTACTTAGGA AGACCACAC TAGATGGACA AACAGCATTG GGAGGCTTA GCGTGCTCC 3060
 TCTERATTC ATCCCTABA GAACAGGAST CAGGAGCCGC TGGCAGGAGA CAGIATGTCA 3120
 CCCAGGACTC TCCCGGTGCA GAATATGAAC AAYGCCATGT TCTTGAGAA AACGCTTAGC 3180
 10 CTGAJTTTCA TAGGAGTAA TCACCAGACA ACTGAGAAT GTRGACACT GAGCAGGACA 3240
 GGTGACCTGT CTCCTTACA TAGTCCATCT CACCACAAAT CACACAACAA AAATAGARG 3300
 15 AGATATTTTG GGTTCAAAAA AAGTAAAAAG ATAATCTAGC TGCATTTCTT TAGITATTIT 3360
 GACCCCCAAA TATTTCTCA TCTTTTCTT GTTGTCATG ATGGTGCTTA CATGACTTG 3420
 TTTATAGAGG ACAGGTCAGC TGCTGGTC AGTSATCTAC ATTCTGAAGT TGTCTGAAA 3480
 20 TGTCTTCATG ATTAAATCA GGTAAAGCT TTTCCGGGA AACTGCGAGA GACAATGCTG 3540
 TGAGTTTCCA ACCTYAGCCC ATCTGGGSC AGAGAAGGTC TAGTTTGTCC ATCASCATTA 3600
 25 TCATGATATC ACCACTGCTT ACTTGGTTAA GGAGGGTCT AGGAGATCTG TCCCTTTTAG 3660
 AGACACCTTA CTTATAATCA AGTATTTGGG AGGCTGCTT TCAAAATAG AAATGCTCTG 3720
 TATTCRATG ATCATCTCT AAACATTTTA TCATTTATTA ATCATCCCTG CCTGTGTCTA 3780
 30 TTATTATATT CATATCTTA CGCTGGAAAC TTTCTGCTC AATGTTTACT GTGCTTGTG 3840
 TTTTGCTAGT GTGTGTTCTT GAAAAAATA ACAITCTCTG CCTGAGTTT AATTTTGTG 3900
 35 CAAAGTTATT TTAATCTATA CAATTAAGG CTTTGTCTA CCAAAAAA AAAAAAATA 3960
 AAAAAAATA AAAAAGCGGA CCCGTGGG 3989

40

(2) INFORMATION FOR SEQ ID NO: 29:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3735 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

CTGCTGTTG CTGGTGGG TCCGACGAG GCTTGGGAG CGGTGACGG GTGCGGGG 60

CTGCTGTTG CTGGTGGG TCCGACGAG GCTTGGGAG CGGTGACGG GTGCGGGG 120

60

CTGCTGTTG CTGGTGGG TCCGACGAG GCTTGGGAG CGGTGACGG GTGCGGGG 180

	ATTTTACTG GCAAAGAAAT CCGGGGAGAA TGTGGCCAAG TTTATTATTA ATTGATACCC	360
5	CAAATATTTT CAGAAGGACA TAGCTGAACC TCATATACCG TGTTTAATGC CTGAGTACTT	420
	TGAACCTCAG ATCAAAGACA TAASTGAAGC CGCCCTGAAG GAACGAATTG AGCTCAGAAA	480
	AGTCAAAGCC TCTGTGACACA TGTTTSATCA GCTTTTGCAA GCAGGAACCA CTGTGTCTCT	540
10	TSAAACAACA AATAGTCTCT TGGATTTWTT GTGTTACTAT GSTGACCAGG AGCCCTCAAC	600
	TSATTACCAT TTTLAACAAA CTGACAGTC ASAAGCATTG GAAGAGGAAA ATGATGAGAC	660
15	ATCTAGGAGG AAAGCTGTC ATCAGTTTGG AGTTACATG CGAGCAAAAA ACAAGCTGA	720
	GAGAACTTTT TCTTAATGC CAGAGAAAAA TSAACATTCC TATTGCACAA TGATCCGAGG	780
	AATGGTGAAG CACCGAGCTT ATGAGCAGGC ATTAACTTG TACACTGAGT TACTAAACAA	840
20	CAGACTCCAT GCTGATGTAT ACACATTTAA TGCATTGATT GAAGCAACAG TATGTGCGAT	900
	AAATGAGAAA TTGAGGAAA AATGGAGTAA AATACTGGAG CTGCTAAGAC ACATGGTTGC	960
	ACAGAAGGTG AAACCAAATC TTCAGACTTT TAATACCATT CTGAAATGTC TCCGAAGATT	1020
25	TCATGTGTTT GCAAGATCGC CAGCCTTACA GGTTTTACGT GAAATGAAAG CCATTGGAAT	1080
	AGAACCCTCG CTTCACACAT ATCACCATAT TATTGCGCTG TTTGATCAAC CTGGAGACCC	1140
30	TTTAAAGAGA TCATCCTTCA TCATTTATGA TATAATGAAT GAATTAATGG GAAAGAGATT	1200
	TTCTCCAAAG GACCCGGATG ATGATAAGTT TTTTCAGTCA GCCATGAGCA TATGCTCATC	1260
35	TCTCAGAGAT CTAGAACTTG CCTACCAAGT ACATGGCCTT TTA AAAACCG GAGACAAC TG	1320
	GAAATTCATT GGACCTGATC AACATCGTAA TTCTATTAT TCCAAGTTCT TCGATTTGAT	1380
	TTGTCTAATG GAACAAATTG ATGTTACCTT GAAGTGGTAT GAGGACCTGA TACCTTCAGC	1440
40	CTACTTTCCC CACTCCCCAA CAATGATACA TCTTCTCCAA GCATTGGATG TGGCCAATCG	1500
	GCTAGAAGTG ATTCCTAAAA TTGGAAGA TAGTAAAGAA TATGGTCATA CTTTCCGCAG	1560
45	TGACCTGAGA GAAGAGATCC TGATGCTCAT GGCAAGGGAC AAGCACCCAC CAGAGCTTCA	1620
	GGTGGCATT TCTGACTGTG CTGCTGATAT CAAATCTGCG TATGAAAGCC AACCCATCAG	1680
	ACAGACTGCT CAGGATTGGC CAGCCACCTC TCTCAACTGT ATAGCTATCC TCTTTTAAAG	1740
50	GGCTGGGAGA ACTCAGGAAG CCTGGAAAAT GTTGGGGCTT TTCAGGAAGC ATAATAAGAT	1800
	TCCTAGAAAT GAGTTGCTGA ATGAGCTTAT GGACAGTGCA AAAGTGTCTA ACAGCCCTTC	1860
55	CCAGGCCATT GAAGTAGTAG AGCTGCAAG TGCCTTCAGC TTACCTATTT GTGAGGGCCT	1920
	CACCCAGAGA STAATGAGTG ATTTTGCAAT CAACAGGAA CAAAAGGAAG CCTAAGTAA	1980
	TCTAACTGCA TTGACCACTG ACAGTGATAC TGACAGCAGC AGTGACAGCG ACACTGACAC	2040
60	CAGTGAAGGC AAATGAAAGT GGAGATTCAG GAGCAGCAAT GGTCTCACCA TAGCTGCTGG	2100

	AATCACACCT GAGAACTGAG ATATACCAAT ATTTAACATT GTTACAAAGA AGAAAAGATA	2160
	CAGATTGGST GAATTTGTTA CTGTGAGGTA CAGTCAGTAC ACAGCTGACT TATGTAGATT	2220
5	TAAGETGCTA ATATGCTACT TAACCATCTA TTAATGCACC ATTAAAGGCT TAGCATTTAA	2280
	GTAGCAACAT TGCGGTTTTC AGACACATGG TGAAGTCCAT GGCTGTTGTC ATCAGGATAA	2340
10	GGCTGCACAC CTAGAGTGTG GGTGAGCTGA CTTACGAGT CTGTGCTGCT GGTATTGCCC	2400
	TCTGCTGCTG CTGGACTTCT GCGTTTGTG GCTCTAIGTG CTGCTGATAT GGTGGTCTTT	2460
	CATCTTAGGT GTTCATGAG TTCTAACACA GGTGGGGTTG GGTCAATAGT TTCCCAATTT	2520
15	CAGGATAATTT CGATGTGAGA AATAAGCAT CTTAGGAATG ACTAAACAAG ATAATGAGAG	2580
	TTTAGGCTGC ACAACTGTA AATGACTGT AGATAAATGT TGTAAATAGT GTACAGTTT	2640
20	GTTTTTTTBT TAATATAGCC CCGCCATAG TTTTCTAGT TGAACAGGCA TGAATGTTTC	2700
	ATGCTCCCT TTTTTCTTG TCTATAGCTG TTACCTAATT TAGTGCTTGA AATGAGAGCT	2760
	AGTGATGACA GAAGGATGTG GAATGTCTTC TTGACATCAT TGTGTATTCG TGGTAATCAA	2820
25	GTTGGTAACG ACTACTTCTA GCACTCTTTA CCACTATGAC TTAAGTGCTC CTGGAACGCA	2880
	GTAAGTGGAG GTTTCAGCA TTCTGCCTT CATGAGGGCT TCTACCACTG ACCACTTTTC	2940
30	ACGTACCTGG CTCCAGATT TACTTAGGTA CCCCAGAGT CGTCCACATA AGCAGCTTCA	3000
	TCTTTACCTT GCCAGAGTTG ACAATTATGG GATACTCTAG TCTACTTATA CTTGTGTTCC	3060
	CATCTGCTTG CCATCCTCTG AAGGCTAGCA CCCAGTATA CATCCTTAGA AACCAGAGTA	3120
35	TGGTTTTTGT TTTCTCTGG AATGTAGCT CTTAAGGAT TTAATTGAGG GACAAAAAA	3180
	AAAAAAAGCC GATATAGTAG CTAGCTACTT AAGCATCCAT GGGTATTGCT CCATATCAAA	3240
40	GGAGATTTCG AGGACAGAAA GAGTAAATTA GCTTTGASTC TTGGTTTACA GCTTCCAAAG	3300
	AGAGCCTTGG CCAGCTGAAA TGTAACTCG GTTCTTCTCT GTCTCTAGTT CATCAGGAGC	3360
	TGCAGATGCC TGACTCTTCT TAGCTTACT ATTCAATACA GTCTTAGAT TCAGGTATG	3420
45	CTCTTCTTA TCCAGGCAAC TATTTGAAT CAGCATGTTG CTTGGAAGT AGATTGATA	3480
	GGAGAAAATC CATTTGGGTA GATGCCCTAT GAATTTGTAG TAGATTTCA AATGAGTGA	3540
50	TTTGTATGCT TGGTACTTTT AAGTTGTGAG TACAGATCCT CCAAAACCAT ACTCTGAGCA	3600
	ATTAAGTACC TTGAACATAG AGAAATTAAG GGGCTCAGAG GATGAGTCTG CATCTCTCT	3660
	ATTAAGTACC TTGAACATAG AGAAATTAAG GGGCTCAGAG GATGAGTCTG CATCTCTCT	3720

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1657 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

5 TAGTAATTCA TTAACTCCT CTTACATGAG TAGCGACAAT GAGTCAGATA TCGAAGATGA 60
 AGACTTAAAG TTAGAGCTGC GACGACTAGC AGATAAACAT CTCAAAGAGA TTCAGGACCT 120
 15 GCAGAGTGGC CAGAAACATG AAATGGAATC TTTGTATACC AACTGAGCA AGGTGCCCCC 180
 TGCTGTATAT ATTCCGCCAG CTGCTGCTCT TTCAGGGAGA AGACGACGAC CCACTAAAAG 240
 CAAAGGCAGC AAATCTAGTC GAAGCAATTC CTTGGGAAT AAAAGCCCCC AGCTTTCAGG 300
 20 TAACTGTCTT GGTCAAGATG CAGTTTCAAT CTTGCACCCC CAGCAGACCC TCCACCTCTC 360
 TGCCAACATC CCAGATTCGG GGCAGAAATC GCTGTACAG CCCCTTAAGC CATCTCCCTC 420
 25 CAGTCACAAC CTCTATTCAG CCTTCAACAG TGATGGTGGC ATTTCACTAC CAAGCCTTTC 480
 TGCTTCAGST CAAAGCAATC GAGGCAAAA CACTGTTGGG GCAACAGTGA ACAGGCAAGC 540
 CGGCCAAGCT CAGCTCTCTG CCATGACCTC CAGCAGGAAG GGCACATTC CAGATGACTT 600
 30 GCACAAGTTG GTAGACAATT GGCCTCCAGA TGCCATGAAT CTCTCAGGCA GGAGAGGAAG 660
 CAAAGGCCAC ATGAATTATG AGGCCCCCTG AATGGCAAGG AAGTTCTCTG CACCTGGGCA 720
 35 ACTGTGCATC TCCATGACCT CCAACCTGGG TGGCTCTGCC CCCATCTCTG CAGCATCAGC 780
 TACTCTCTCA GGTCACTTCA CCAAGTCTAT GTGCCCCCCA CAGCAGTATG CCTTTCAGC 840
 TACCCCATTT GCGCTCAAT GGAGTGGGAC GGGTGGGCA CCACCACAGC CACTTGGCCA 900
 40 GTTCAACCT GTGGGAATG CCTCCTTGCA GAATTTCAAC ATCAGCAATT TGCAGAAATC 960
 CATCAGCAAC CCCCAGGCT CCAACCTGCG GACCACTTAG ACCTAGAGAC ATTAAGTGAA 1020
 45 TAGATCTGGG GGCAGGAGAT CGAATGCTGA GGGGTGGGT GGGGTGGGA AGTAGCCTAT 1080
 ATACTAACTA CTAGTGCTGC ATTTAACTGG TTATTTCTTG CCAGAGGGGA ATGTTTATAA 1140
 TACTGCATTG AGCCCTCAGA ATGAGAGATC TCCCCGCTC CAGTTATTGG AATGGGAGAG 1200
 50 GAAGGAAAGA ACAGCTTTTT TGTCAAGGGG CAGCTTCAGA CCATGCTTTC CTGTTTATCT 1260
 ATACTCAGTA ATGAGGATGA GGGCTAGGAA AGTCTTSTTC ATAAGGAGCC TGGAGAACTC 1320
 55 AATGTAAAT CAAACCCATC TGTAATTCG AGTGGGTGGA GCTCTTGCTT TTGCTACATG 1380
 CCTGCAATCC CTCACTCCCT CAAGAATCCG AACCACAGGA CAAAACCCAC CTACTGGGCT 1440
 60 CTCTCCTACC CTGCTCTCCT CCGTTTTTTT TACCCCTCTC TTTTTATTT TTCTTTGCT 1500

CTTTAGAACC CAGTGA AAAA TACCAGGGTA CTGGGGTGCA ACTCTTTCTT ATGATAGGTC 1560
 ATTAGTGCTT TAAGCAAAAAG ATATTAGCAG CTTTGA CTGC AGCATTAGCA ATTAGGAAAA 1620
 5 AAAAAAANWA AAAACTCGAG GGGGGGCCCG GTTACCAAT TCGCCCT 1667

10 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1408 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

20 ATTACACACC TGAGCACTGT GCTTGGCAAG ACCTGTCTTA ATAGATTAGA GAACCACTGA 60
 TAGATGCTCA GCTTTCTGTA GCACTGAGAA CCTACATTT CAAATGTGGA TAGCACCTTT 120
 GGGGGGAAAG ATCACTTGGC AATCTGCAAT TCTTTTTTGA CACAGGGTCT CACTGTGTTG 180
 25 CCCAGGCTAG AGTCATGGC AGGATCTTAG CTCACTGCAA CCTACACCTC CCAAGTTCAA 240
 GGGATTCTTC TGGCTCAGCC TCTTGAGCAG CTGGGATCAC AGACATGGGC TACCATGCCC 300
 30 AGCTAATTTT TTATATTTT TGTGTGTTT TTTTCTTTK TAAGTAGAGA CGGGCTTTCA 360
 CCACGTGGGS CAGGCAGGTC TCGAACTCTT GATCTCAGGT GATCCACCCA CATCTGCGTT 420
 CCAATATCTT TCTCAACATA ATGATAGGCG TAATTAATAT TTTCCAGTAC ATTTTATGCG 480
 35 CTTTACACAC GAGAGTGGTA GACAGACACA AACCAGATC TGTCTGACTC CAAAGCCCGT 540
 TTGTCACTAT TCGTTTACG GTATCTCTATA GTGGTATCTT TTACAGAAAG ACAGCTTTTA 500
 40 TCCAACAAAG ACTTAACTTC CCAGGATGCC AGAAGGACAA AGCGGGATTG CTTTAAAGRA 560
 GFAAGTTATC AAGAMCTTAT TTTATAAATG AGATTAGATA GGGAAAGGCA ATTTATCTTT 620
 45 ATTAAAAACT GAAAAGGCGA GATAGGGGAA GGAGTCTCTT CGTTGGTTT TTTGAGGGAA 680
 ATACTTCAT TCGTTTATAT AAAAAAGAT AGTACCTAAG GTTTGAGGT AGGWACAGCT 740
 TAAGGCATGC TAATGCTCAT GGGTCTTCC ATAGTCATTT TGTATTTTG GTTACATTT 800
 50 GAGCAATAGG CAGCCCTTCA CTGCTGCTGG AYTCACTCTT GCAATATTA CAGGTGACAG 860
 AGGAGACAGG AGGTATGTT TTTCTATTTT TAWACATGCT TTATATTTAA CACAAGCTCT 920
 TGGTATCTT AGATAACAG AAGTTGCTTA CCACTCTTT TACTGATTT AAGCTTTTAA 980
 60 TCTTAAGAA ATGATGCTT TATGTGTA GCTTATAT AAGGATTT TATAAGAA 1040

	AGATTGAGAC GTGCTTCAG AATAATGCA TTCTTTTGCA AAGGTGAATA TTTTCTCTT	1320
5	AAAAATATG TATTAAGTGG TATGTCAT TATTAGTCTT CCTAAAAAA AAAAAAAAAA	1380
	ACTTGGAGGG GGGGTCGGT ATCCATT	1408
10	(2) INFORMATION FOR SEQ ID NO: 32:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 2031 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
20	(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 32:	
	AGGATATGCA TGATTCTTAA CAGGCTATA TGTAAAAAA AAATTGGAAA ATCCAATACA	60
	TTTTTACTTA TACAACTAC AGAATCAGTA TGCAAGTTTT ATTTATCAAA ATGTAATGGA	120
25	TTTTTAAGG CTGAGAAAT TTGCTTATAC CTACCTTTTC AGTTATTTTA ATTATACCAA	180
	ATTATCAACT AGAATAGCTT CATCCATATG AAATATAAAA TGAAGAGACA CCTAGGCTCT	240
30	ATCAGGCTTA GGATTCTTTG AACTTATTTT CACTTTAATT TCTCAGTGGA AGTTAAGAGG	300
	GGTGAGAAA CAAAGAGGG GAAAACTGA CAACTAACAA AACCAGCACC ACATCGCTAG	360
	GTGGTGCTTA CTAATTACCT TCTCAGGATT TTCCTCAGAT TGAAAAGCTT ATGAGGATTT	420
35	CTGGGAGTC TTAATAGCT CCTGTAGT ACAGAGCTTT CCGTAGATA TTACTCTTG	480
	AGCACATGTG GTTGTAAC CTTAATTTC TTTCTCCAGG AGGGTGGTGA TAGAAACAGA	540
40	TGGTAGTATT TATGAAGTGA TTTTCTCTG AAATGTTGAG CGTGGGGAGA AAAGACTTTA	600
	AGGGAGGAGA GCCATCTATT TTGTTCTTAA AGCCACCTCT CAGCAGAATC GTCATGTTTT	660
	TCTGATGCAC CGCTCTGCTT CATGCCAAG ATGACTTGCG AGGCAATCTC AGGAGCTGTG	720
45	GACTTAACCT TTGCAAAGCA CACTGTCTTT CTCAGCGTTC TCTGCAAGTC AGTAGGTGTT	780
	AGTATGGTTG CAAAGTTCAC TGTCTCAGCA AAGTTGAAC TGGGTACCTC TCTACAGCTG	840
50	TTTCTCAGA GGGAAAAATC TTGAGACCAG ATGGTGGAGC TCTGGAGTCA GAGGAAATGG	900
	GTGTCTTCAG CACAAAGCTG CTGCTTTTAC TTCAGCCACT TCTGACATTT TTACATACCG	960
	AGCCTGAGAT TRTGTGATTA TCTCAATCA AATCACTTTG ATGGAGATAA ATAATCAAAA	1020
55	CTGTTTCTATA GTCATTGATT TGGTGAGAAC AGTAATGGAA AATGTTGTTG AAGGACTTCT	1080
	CATTTTGTGA GCTTTCTTTC CAGAGTCTTG GCTGATTGGT GTTCGCTGTT CATCTGAGCC	1140
60	CCCCAAAGCA TTATTACTGA TACTTGACA CAGTCAAAAG CGCAGACTGG ATGGATGGTC	1200

TTTTATAAGG CATTTAAGGG TACACTACTG TGTTTCACTG ACCATACATT TTTCTTAGCC 1260
 CCTCAAGTAA TATAGCAGAG AGTTATGAAT GACAATTGCG CTAACCATTC CTCTTCATAT 1320
 5 CTGCGCTCTTC CCTTACCAT CGTAATTCTC CAAACTGCTC ATAAAGGCAC TCTGTGAAGA 1380
 TATTGGGGAG TACATCTTA AGTCTCACC TGCTTGCAT AGGAAAGGCC AACTGACGA 1440
 CAAAAAAAAA ATTCTTTATA AAGATGATAT GGTAACTGT ATTTTTGCCC TGGTCTGGG 1500
 10 TGGGTCCAGT CAGTCTAGA TTTACAAGCA TTTAGGAGCC TAGGTAAAG CTGCTAGTAT 1560
 TCTTTTAAAA GTTACATTTA TGACTTGCAA TCATAGAAAA CTCTTCCAA TTAAATGGCA 1620
 15 TTTTATAATA TTATGTGTGT ACTTCACAGT GTTAAAAATA CCTTCATAG TTATTCATT 1680
 TCACTCTTCA AGAAAGTCCA TTTTAACCAG TACTCTGGGT GCAATAAATA ATATGTAGAA 1740
 ATTAAAGTCC TCGAATTCCA GCATATCCAG TGAGTTTCTA CAGTGTGTTT ATGTGGAATG 1800
 20 TTTAAGGATA TACAATTGTA CTTTATATAA ATTGSTTTCT GTTCTTCTTA AATGTACAT 1860
 GAAATAATTC TGCTGTACA TTACTCTGA AATTAACAGG GAAAGGGGA AGAGCTCTTG 1920
 25 GCTCCCTTGA GGTCTGTCTA GTGTGTTAG CAGTGTGTA AACTGAGCTT TTAGTAACCA 1980
 TTTAACGTA TGTAAACTTG GTTTCTAATT AAAAAAAAAA TTTCTTTCC A 2031

30

(2) INFORMATION FOR SEQ ID NO: 33:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 971 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCGTCGGAA CTGGGCGCG GGACATCCAC GGGGCGGAG TACACGCGG GAGGAGAGC 60
 AGTGTCTGTC TGAGGCGGAT GCGAAAAACC ATGCAATTTCT TATTGAGATT CATTSTTTTC 120
 45 TTTTATCTGT GGGGCTTTT TACTGCTCAG AGACAAAAGA AGGAGGAGAG CACCGAAGAA 180
 GTGAAAATAG AATTTTCTCA TCTCCAGAA AACTGCTCTA AGACAAGCAA GAAGGGAGAC 240
 50 CTACTAAATG CCGATTATGA CCGTACCTG GCTAAAGAG GTCGAAATT CTACTGCAGC 300
 CGGACACAAA ATGAAGGCCA CCGCAATGG TTTSTCTTIG GTGTGGGCA AGTCATAAAA 360
 GCGCTAGACA TTGTATGAC AGATATCTCT CTTGAGAAA AGGAAAAGT AGTTATACCT 420

60 AAAAATAA AGAAGGAGAA TATAGGCA CTCTTTAAA GAGATAAA CTCTTACTTG 480

CAAAGGGAAT TTGAAAAAGA TGAGAAGCCA CTTGACAACT CATATCAGGA TGTAGTTTAA 655
 GAAGATATTT TTAAGAAGAA TACCATGAT GTGATGGCT TCATTCTCC CAAGGAATAC 720
 5 AATGTATACC AACACCATGA ACTATAGCAT ATTGTATTT CTACTTTTTT TTTTAGCTA 780
 TTTACTGTAC TTATGTATA AAACAAAGTC AATTTCTCC AAGTTGTATT TGCTATTTTT 840
 10 CCCCATGAG AAGATATTTT GATCTCCCA ATACATTGAT TTTGGTATAA TAAATGTGAG 900
 GCTGTTTTGC AAACITAAA AAAAAWAAA AAAACTSGAG GGGGGCCCGT ACCCAANTCG 960
 CCGNATATGA T 971
 15

(2) INFORMATION FOR SEQ ID NO: 34:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1792 base pairs

(B) TYPE nucleic acid

(C) STRANDEDNESS: double

25

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAACCCCTT TCTCTGGTA AAGGGTAAG GCGGGATAA TGTITACCAC AGGTACGAAA 60
 30 TAGTCACTTT AACATTGAGA CCTCTGCCCTC ATTGAATTCA GGTTTTTTAA GTACTTGAAA 120
 CTCTTCAGAT TCTCTTATT TTAGTTTCTT TTACATTTA TGAAGTAGAA AGCATTGTTT 180
 35 TGTAAACTGT TTTGAAAATA AATAGCCTAG TCTCTTATCC TCTTTAGCGT GGATTAAAGG 240
 TGAAGTTCTG CAAATGGGAG AGTGTTCACA GTAGATAGCT CAGATTGATT GAACACATTT 300
 GAGGAAGAGA CTCTTCATG AGATACCAGC ATTTTACAA ATACTTTTTA TGTACATTCT 360
 40 TTATTTGTC ATTTTGTCAA CCTCTCCC CAGCACATCT TCTTCTTTT TACTATGTCT 420
 ATGTAGGGAA AAACAAAACA AAAAATTGCA CTTACGTTAC ACTCCCAAAA TGTGGGTAAT 480
 45 CCGTGTCTTT CAAAAACAT TTCTGTTTTT TGTTTGTTT TGTTCAGTCC ATTGCATAAG 540
 TGACAAGTTT GGTGCTTGT GGCACGTATG TATGAAGCG GAGGGGATG ASAATTGCCT 600
 GTCCTTCAGT ARGCTGTAA AGTAATTTAC ATGTAAGTAA AAAGGAAAA TAGAATAGAT 660
 50 GCCAAAGTCA TTATTTCAGT CCTTAGTTTT CTATGTGGC ATTACTGCAT CTGCTAGTTA 720
 GTGAGAAAGC ACCCTCAGCT TTTACTGCTC CCTCCCTGC CTGCCAACAC ACTTGATGTG 780
 55 TGCAAACAGC CTTCAAGTAT CTGTGAGATG ACCTATATAA GGTATTGAAT AAGGTATTCT 840
 TGTCAATTAA GAAATGGACT GGATAAAACT TACTTGGTTG TCATTATTTT ATCTCATTTG 900
 60 TCCTGTTACA TGCCCTATGT TAAGATAATT ATATTGCCAC TAATAATCAA GATGCTAAAT 960

GAGTATTACA ACTGGCTAAT ATCAATTTTT ATATACAAGG GTATGTGTAT ATTTGGAATT 1020
 GRTATGAGAA ACTCATTTGT ACCCATTTGA GTGATATTGC ACAACAAACA CAGATAYCTA 1080
 5 CAGACTCGGT TTTCATTTTC TCGTGTCTT TATGATAATG ATCTTTGTAG ATTTGGTTATT 1140
 TCTGTACTTT ATCTSTAATA AATTTGTAG ATCTGTGAA CCATTACTTT GCTTAAATCA 1200
 10 TTGAGACTT GAGCTTTTAA TAACAAAGCA TCAATATTCA CTAAAGTCAA TCTCTTTTGA 1260
 GTTCTGTGA CTTGGCTAGA AGCTCTGAC ACTAAGGAT TAGTGTAAAT TTTCCTGCG 1320
 GGTCTTAC TAGGGCATT CTSTATAATG ACTTGATTT GGCACATAGA CTTCAGATA 1380
 15 TATAATATTT TGAGTATTT GTTGATGCG CTATSTTTTA TTGCATAGTG TGAAACGTGT 1440
 AAAGCTTGT TAAGTGTAT ATAGATAGCT TATTGTGAC TAGTTATAGT GTATTTAGCG 1500
 TTGCTGTAA TATTTAAGCT TCTTTACTGA TGTGTGTGT GGTAGGAACA TATAATTTTT 1560
 20 GTACATTATA TTTACTGAGA TGTGCTTT TTTATTTTAC AAATACTTG GAATTCGAAT 1620
 GTSTTTTTG CTTGCTGAG GATTAATTTG GAAAGTTTT TAATGACATT CCACTGATTT 1680
 25 CAGATTTTG TTGAGATTGA CTTCAATAAA TGTCTCTTA TGTTCAAAA AAAAATTA 1740
 AAAGTGAAG GGGGCGGT ACCCAANNCG CGGATATGA TGTAAACAA TC 1792

30

(2) INFORMATION FOR SEQ ID NO: 35:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 896 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

AGTTGNANAC AACAGGACCT GATCTCTTGG GCAGCACCAG TAGGTTGCCC TTTGCTCTT 60
 GCGAGCTTCA CTTGCGACTT TTTGCGCTT TGGGATGCG TTGGCAGACA GAGTCTTTG 120
 45 TTGCTGTGAG TGGGATCTT TGGCTTGG TTTCTGTGAG CTTGCGCTT CTTTCTTGT 180
 CCGGGGACG CTTGTGTGAC TTGGCTTTT CCGTCTTTC CTTTCCAGGA CAGGACGGC 240
 50 GAGGAGGTGC GAAAAAACA GAGCTGAAG GAAGAGGCT CCAGGTAAAG CCTAGAGGCG 300
 AAAGAACTTT CCAGGTGAC GAGACGCTC GAGCAGTCC ACGTTCCAGG CAGGCTCGMC 360
 CCGGCTTTC CTTGCGACA CTTGCTTTC GGGGAGGCG CTTGCGACA GAGGCTTTC 420
 60 AGAAGTAAAG GAGATTTTAA CCGTAAATG CAGATTTT ATTTGAGAT TTCTAGCTG 480

ACATTGAGCC TCCCAGACAC CATGTTGAGG AGAGATGAAA ACCAGGGCGG TAGAACTTCA 660
 GGGTGAAGGA CAGAGTCTCT GGTGGGGCAG CGGCTGCAGG GCGCACCAGA GAACCCAGCC 720
 5 AGAGGGGCTG TGAATACCAG TGGTGTGCT TCCACCTGCG AGCAGGTGGG ATGAGGTCTG 780
 TGTGTGTCTG TGAACCATCA TTTTGTGATC ATCATGACCA ATGAAACATT GAAAAAAAAA 840
 10 AAAAAAATG GAGGGGGGCC CGTACCCAAN TCGCCGNATA GTGATCGTAA ACAATC 896

15 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 912 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

25 TCGACCCACG CGTCGGTCA GCCAGTGGCA TCCAGCCATG ACAGCCTTCT GTTCCCTGCT 60
 CCTGCAAGCG CAGAGCTTCC TACCCAGGAC CATEGGCAGG CCCCAGGACA GCTTCAGACC 120
 AGGGGAGGAA GACGAAGGGA TGCAGCTGCT ACAGACAAAG GACTCCATGG CCAAGGGAGC 180
 30 TAGGCCCCGG GCCAKCGGCG GCAGGGCTCG CTGCGGTCTG GCTACACGC TGCTGCACAA 240
 CCCAACCTTG CAGGTCTTCC GCAAGACGGC CCTGTTGGGT GCGAATGGTG CCCAGCCCTG 300
 35 ARGGCAGGGA AKGTCAACCC ACCTGCCCAT CTGTGCTGAG GCATGTTCTT GCTACGATC 360
 CTCCTCCCTC CCGCGCTCTC CTCGAGCAT CACACCAGCC ATGCAGCCAG CAGGTCTCTC 420
 GGATCACYST GGTTKGGTGG AGGTCTGTCT GCACTGGGAG CTTGARGAG GCTCTGCTCC 480
 40 ACCCACTTGG CTATGGGAGA GCCAGCAGGG GTTGTGGAGA AAAAACTG3 TGGGTTAGGG 540
 CTTTGGTCCA GAGCCAGTT GAGCCAGGCG AGCCACATCC AGGCTCTCC CTACCCCTGG 600
 45 TGTGCCATCA GCCTTGAAGG GCCTCGATGA AGCCTTCTCT GGAACCACTC CAGCCCAGCT 660
 CCACCTCAGC CTTGGCCTTC ACGCTGTGGA AGCAGCCAAG GCACTTCCTC ACCCCYTCAG 720
 CGCCACGGAC CTYTYTGCGG AGTGGCCGGA AAGCTCCCG 3CCTYTGCC TGCAGGGCAG 780
 50 CCCAAGTCAT GACTCAGACC AGGTCCACA CTGAGCTGCC CACACTCGAG AGCAGATAT 840
 TTTGTAGTT TTTATKCTT TGGCTATTAT GAAAGAGGTT AGTGTGTTCC CTGCAATAAA 900
 55 CTTGTTCTCTG AG 912

60 (2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1332 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

10 AATTGGGCAC GAGGCGAGGC GAGGGAAACT RAGGCGGAAA GTTGTGTGTC GTGTTGGCAG 60
 GAGGGCCTAG AAGGGAAGA CTGTCTAGTG GGACAACTC ATATTATAAA TTTGGAATGC 120
 TGAATAGAAA ATTATAGATT TTGATATTGA ACGAAATGAA GCGAAGCTA AATGAAAATT 180
 15 CAGCTCGAAG TACAGCAAGC TGTTGCGCTG TTCCGTTTCT CAATCAGAAA AAGAGGAACA 240
 GACAGGCATT AACTTTAAT CCACTTAAAG ATGATTCAGG TATCAGTACC CCTTCTGACA 300
 20 ATTATGATTT TGCTCTCTTA CTTACAGATT GGCCCTGGGA AGCTGTCAAT CCACACTTGG 360
 CTCCTGTAAT GAAACACTG GACACCGGGC AAATACCACA TTCAGTTTCT CGTCTCTGTA 420
 GAAGTCAAGA TTCTGTCTTT AACTGTATTC AATCAAATAC TGAAGAAAC CAGGTGGTTT 480
 25 GGAGCTACAG AGATGTGTAAC AAAAATACCA GCTTGAAAAC TTGGRATAAA AATGATTTTA 540
 AGCCTCAATG TAAAGGAACA AACTTAGTGG CAAATGATG AAAAAATTCT TGTCCAATGA 600
 30 GTTCGGGAGC TCAACAACAA AAACAATTAA GAACACCTGA ACCTCCTAAC TTATCTCGCA 660
 ACAAAGAAAC CGAGCTACTC AGACAAACAC ATTCATCAAA AATATCTGGC TGCACAATGA 720
 GAGGCTAGA CAAAACAGT GCACTACAGA CACTTAAGCC CAATTTTCAA CAAATCAAT 780
 35 ATAAGANACA AATGTTGAT GATATTCAG AAGACAACAC CCTGAAGGAA ACCTCATTTG 840
 ATCAGTTACA GTTTAAGGAA AAAGCTAGTT CTTTAAGAAT TATTTCTGCA GTTATTGAAA 900
 40 GCATGAAGTA TTGGCGTGA CATGCACAGA AACTGTACT TCTTTTTGAA GTATTAGCTG 960
 TTCTTGATTC AGCTGTTACA CCTGCCCCAT ATTATTGAA GACTTTTCTT ATGAGGGATG 1020
 GGAAAAATAC TCTGCTTCT GTCTTTTATG AAATCGATG TGAACCTCC AGACTGATTA 1080
 45 GAGGCGGAGT TATAGATGT GTTGGCAAT ATGACCAGAA AAAGAACATT TTCCAATGTG 1140
 TTTCTGTGAG ACCGGCTCT GTTCTGAGC AAAAACTTT CAGGCATTT GTGAAAATTG 1200
 50 CAGATGTTGA GATCAATAT TATATTAATG TGATGAATGA AACTTAAGTA GTGATAAAG 1260
 GAAGTTTAGC ATAAATTATA GCACTTTTCT GTTATTGCTT AATTACCAT CTCATAGTT 1320
 TTTTCTGCTGAG ACCGGCTCT GTTCTGAGC AAAAACTTT CAGGCATTT GTGAAAATTG 1380

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

GGGCTACTTC AAAGCCCTGG GCGTTATTTT TTCAGGTAAA AAAATATAAA GTCAGATCTC 60
 ATGCGGGGCTG GCGATGCTGT TAGACCCCTTT CATCTTTCTC TTCTGCCTCT TGTCAACAGC 120
 TGCCAGTCC TCTTTGGAAT TCATATACAT ACAGTTCTAA TACTGATGTA TTTACCCCTCA 180
 TAAGCCACTG AACCCAGAAT CTTATTTGAA TTATAATCCA GAAACATCAG GTGACGTGTG 240
 AGACTACTGT ATGAGAAAGA GACAGTTTAA GGGTCAGTCC AATGGAAAAA AGAGTTCTCA 300
 GAGCTTTCTT TAGCTTATTC TCATCAAAGA GCTTTCTCTG CAGAAGGAAC CTACTGGTTC 360
 CTGCTTTTCA GTCTAGAAA TCTGACCTA GAGTGGCTTA ATCCTGCTAG CACCTCTCTC 420
 TGGTACTCTG GTGCCAAATG ACTCCAGGAA CTGGGCCATG ATGTGGTGGG AATGACCTTA 480
 CCTGAGCAT CTCACTCATG CATTGAACAA CAGCTAAGAG CAGAGCTTAC AGCTTAGAGC 540
 TGGGCCCTGT AAGGTGAGAG GAATCACATC CTGCAGAAGT CTGTCTGTAG AAGCAGGTAC 600
 TCCTGTACAA GCAGAGACAC AGTGGATACC TGAGTAACAA TAATACAAGA CAGGACGTGG 660
 CMACAGCAAA AGATTGCGGT GTCAGAAGAR CCGGAGAACA CTTTCAGGCA GGAACATTCA 720
 PARTTCTTCT TGGAGGAAPT AGGCMCSAAG GGTGGGCAGG ATTTTCMGCG GCAGAGATGG 780
 AGCAAGCAAT TGAATGAAA GCCATGGCAT GGGAAAAGGA GCACTGGCCA CAGGGAGTGC 840
 AACGTGTGTA TGCAAGGCCA CTGTGGAGCC AT 872

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 812 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

55
 60

GGCAGAGGCT CACCCACAGCA GAGATTGAGG GGGAAACCGT ATGAAATTTT TAAGTATTCT 60
 GCTTGATGAT AATAATTTTY CTCTTATGTT AATGTTGGCT CCGTTTGGGT GTTTAGCTTT 120
 TGAAAGGAGT ATGAAATGC GGAATGGGGC TTTGGGCTT GAGGAGCTGT GATCTCTACT 180
 GTTTAAAAAA TTAAATTGCA CAAATAGAAA TAATTCACCC ACATTATTGA ACCCCACTAA 240

AGCATATCCT TTTTGTCCAT ATTCTTTTCC TGCTGCCCTC GTGTGTACCA TTATTACTCA 300
 GTTGTGATTT GAGCTCCTTC CACTTAAAGT CATTGATAGA TACTTTTGTG TCGTGTTKGA 360
 5 ATATTATTTG AATTTCTATT CTGTGTTTTA CTTAATTACT TTATTATGGA ACCTTTACAC 420
 AGGTCTGCTG TACTGTCTCT TGA AAAAGTC TTATTTGAC CACCATCACT GAGCATATAG 480
 10 CTTTTTCCTT ATTTCTTTGG GATAATTACC CGAAGTGGA ATACCGAATC AAACCTCTGT 540
 TTTCTTTCTT TGCACTATT ATATAAATTG TTTTCCAAAC AAGGCATGTT TACAATAGAC 600
 ATTTTTCAAA ATCTGCGPAT TTTCTCTATT TTGCTCTCTG TATGAGAAAT TCAGCGGGGT 660
 15 GCCAAGTCGT TTTCTCTGTG GCTTGACAGA CAGGCTCTGC AGCCACTCT TGCATAGGAC 720
 TAACTACTAC AAATCATGCT GAGACCGAGC TATTTTCTCT GCTTAGARCC TTTCCAGCCT 780
 20 TGAGTAAGTT TCGCATCTG GAAACNITGN AA 812

25 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1515 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40.

35 AATTGGGCAC GAGGGAAATT CAAGCACTTT TCTTAAAGA AGGGGGAATG GATGCTGAAA 60
 CAACACGTNT CCGACAAAGG GAGCAGACAC TGGGCTTGTG AAGCTGCCCC ATACCTTCCC 120
 CACAGAAGTG GGTCCGGCC TCCCTGACAT GCAGATTTC ACCAGAAGA CAGAGAAGGA 180
 40 GCCAGTGGTC ATGGAATGGG CTGGGTCAA AAGCTGGGTG CCTGGGAGCT GAGGCAGCCA 240
 CCGTTTCAAG CTGGCAGCC CTCTGAGCC GAGGTTGGA CCTACTGTG ACACACCTAC 300
 45 CATGCGGACA CTCTTCAAC TCTCTTCTT TCGGCTGGCC TGAGGCTCTG TTCACACTAC 360
 CCGTCAAAG TCAGATGCCA AAAAGGCGG CTCAAAGAG CTCTGAGAGA AGACTCACTT 420
 TTCAGATAAG CCGGTCAAG ACCGGGTTT GGTGGTGAC GACCTCAAAG CTGAGAGTGT 480
 50 GGTCTTGAG CATGCGAGCT ACTGCTGGC AAAGGCCCGG GACAGACACT TTGCTGGGGA 540
 TGTACTGGGC TATGCACTC CATGSAACAG CCAAGGCTAC GATGTCACCA AGTCTTTTG 600

60 TCTAGAGG CTGCAATAG TATGCTCTT CTCTTTAG GATGCACTT ATATATTT 660

300

	CTGGAAAGCTC TTAGACAGT3 AGGATGAGAT AGAAGAGCTG AGAAGACCG TACTCCAGGT	840
	GGCAAAGAAC CAGCATTTGG ATGGCTTGGT GGTGGAGGTG TGGAAACAGG TGGTAGGGA	900
5	GAAGCGGCTG ACCGACCAAG TGGGATGTT CACGCACAAG GAGTTTGAGC AGCTGGCCCG	960
	GGTGTGATAT GGTTCAGGC TCATGACCTA CGACTACTCT ACAGCCGATC AGCTGGCCCG	1020
10	TAATGCAACC CTGTCTGGG TTGAGGCTG GGTCCAGGTG CTGGACCCGA ACTCCAACTG	1080
	GGGAAGCAAA ATCTCTGAG GGTCAACTT CTATGGTATG GATAGCGGA CCTCCAAAGG	1140
	TCCCGCTAG CCTTTCTG GGGCTAGCTA CATCCAGACA CTGAAGGAGC ACAGCCCGCG	1200
15	GATGCTGTG GACAGCTAGG YCTCAGAGCA TTTCTTCGAG TACAAGAAAG CCGCAGCTG	1260
	GAGGCAGCTC GTCTTCTACC CAACCTGAA GTCCCTGCAG GTGCGGCTGG AGCTGGCCCG	1320
20	GGAGCTGGG GTTGGGCTCT CTATCTGGA GCTGGGCCAG GGGCTGCACT ACTTCTACGA	1380
	CCTCTCTAG GTGGCATTC GGGCTCGCG GGTGGACGTG TTTCTTTCTA AGCATGGAG	1440
	TGAGTGAGCA GGTGTGAAT ACAGGCTTC ACTCCGTAA AAAAAAAAAA AAAAAAAAAA	1500
25	AAAAAAAAA AAAAA	1515

30 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 704 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

40	AAGATGGTGG CGCCAGAGC TTCCTCTAT GCTGCTCCCC TGAGAGAGGC GTTTCATCA	60
	ACCAGTTTTC CAAGGAGTTC AATGAGAGGA CAAAGGACAT CAAGGAAGGC ATTCTCTGC	120
45	CTACCAAGAT TTTAGTGAAG CCTCAGAGGA CATTTGAAAT TAAGATTGGA CAGGCACTG	180
	TTTCTACTT CCTGAAGGCA GCAGCTGGGA TTGAAAAGGG GGGCCGGCAA ACAGGGAAG	240
	AGGTGGCAGG CCTGGTACC TTGAAGCATG TGTATGAGAT TCCCGCATC AAAGTCAGG	300
50	ATGAGGCATT TCCCTGCAG GATGTACCC TGTGCTCTGT TGTCCGCTCC ATCATGGGT	360
	CTGCCCCTTC TGTGGCATT CGGTGGTGA AGGACCTCAG TTCAGAAGAG CTGGAGCTT	420
55	TCCAGAAGGA ACGAGCATC TTCTGGGTG CTCAGAAGGA GGCAGATTTC GCTGCCAAG	480
	AAGAAGCTGC CAAGAAGTGA CCTTGCCTC ACCAACTCCC AGATTTCAAA GGAGGTAGTT	540
	GCAAAAGCTG TGCCCAAGGG GAGGAAGGAG GTCACACCAA TATGATGATG GTTTTCATGA	600
60	CTTTGAATGA TATATTTTTC TACATCTAGC TGTATCGAGG CATCAGGCCT GAATAACAT	560

CCTTTCTTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA

704

5

(2) INFORMATION FOR SEQ ID NO: 42

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1094 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	GGCAGCTTTC TTACAAACCC ATCCTTCTGA AATGTTGCTT CAAATTCATC CTCTGCTCCC	60
	CAGTCCCACT ATTCCACACA TACTGTTACT GTTTCCTTAT CCTACTTTCT CAATTTTGA	120
20	ACATAGTTGC AGTTACTGCA TTSAATACCT GTGGGTTTGC CTGTTGTTCT GTCTGTCTCT	180
	GTGTTTCTTG TAATATGGA TCCAGAGAT AAAATGGACA GTTGTNATGC ACAGTTAATT	240
25	CAGAACTAG ACCTTACTTG CTGTGTGAAA TACCAACTAA ATTCTCAGTG AACTCAGCTG	300
	ANCTTTATCT CCTTTTGTTF CCCCAATTTA TAATTTTAST TCAGGCCAG AAAGATGGAA	360
	TCCCAGCTAA GAAATACAAG TTACACCCTG TACTAGCAGC CCATGTGTGC ATGTTCTTTA	420
30	AGTCTCTTTS CAGCTATGTC ATTTATATTG ATTTCCCTST ATTATTATAA GCAAAGCAAA	480
	TTTGAGGAAA AAAACCCATA ATACCACACC TCATTTTTTT CAAGTAATAG GGTTCATAAGT	540
35	CTCATYCTYC ATATAATATG TTGAGTATGC AGTATATTAT GTGTTAGGCT CTGGANAGGC	600
	AGACGTTAGA TCATGTWACA GATQATATEK GATTAGGCAG ATAAACAGTA TTTTAACCTT	660
	TTCCCTATTTA TATGTAACCT GCTTTCAGST TTTTAAATGT TACTATTATG TCTTTAATAT	720
40	ATTATCTTTA TTTGTACTTT TGTATACAGA GTGATTTTCC TTTTTTAAAA AAAATTGTGT	780
	CTTTAGGATG GATTCCAAAG ATGTGCAATC AGTAGGTTTA AGGAATATGC ATATTTTGGC	840
45	TGGCAAGGTS CCTCACACCT GTAATCCCAG CACTTTGCGA GCGTGAGGTS GGTGGATCAG	900
	CTGAAGTCAG GAGTTCCAGA CCAGCCTGAC CAACATGGGG AAACCCCTGTT TTTACTAAAG	960
	ACACACWAA AATTGCCAG TGGTGGTGGC ATGTGCTTGT AGTCCCACTT AGCTACTCGA	1020
50	GAGGCTGAGS CAGGAGAATC CTTTGAACCC GGCAGCCAGA GGTTCAGTG AGGCAAGATG	1080
	GCACCTCTAC ACTC	1094

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1321 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

	TEGCTTAGGC GATCAGCCCTT CCCTTGGCTG GAACTACTGG ACAGACCCCTT TTGAGATGTG	60
10	CCCTGCGGTGG TGTGAGATG TGTAGTAGGG TCTTAGCTCT TTGTTBAGCT TGTGTGTGTG	120
	TTGTGTAGTC TTACCTGTAC GGTGAATTTG GCGGTGTGTT GGAGGCTTTC TTAGCTTTTT	180
15	GGTGAGATTG TATTCCTATG TTTTGTATC ACCTGAATGT TGGTGAAT ATAACTTTGG	240
	TTGTGTAAGG CTCTTTCTTG TGGGAAGTAA GTAGGGGAAA AGGTCTTTGA CGGTTCCTAG	300
	GTCTCTTTGT ACACAGGAA ATCTCTTAA AGCCTTCTT CCAGGAACC TGGGGCTGT	360
20	TGGCAGTGGC TGGTCTGGC CTTTCTGTGT TCTTATCTCA AGGCAGAGCT TCTGAATTTC	420
	AGGCTTTCAT TCGAGAGGCC TTTTGTGGC AGGCTTCTT TTCTGTGAGG AAGGTAGACA	480
25	GGGTGAAGCT GATCTGTAC TTGGGGATC TCTTGGCCT GTTCCAGCAA GTGAGAGAAG	540
	GTACTTACTC TTGTAGCTTC TGTTCAGCCA GGTGCAITAA CAGACCTCCC TACAGCTGTA	600
	CGAAGTACTG TGGCAGAGCT GAGGGAAGGG GATTTCTCAG GTATTTTGA GAACAAGTGC	660
30	TTTAACTACTA GTTAAAGTA GTAACTGCTA CTGTATTTAG TGGGGTGGAA TTCAGAAGAA	720
	ATTTGAAGAC CAGATCATGG GTGGTCTGCA TGTGAATGAA CAGGAATGAG CCGGACAGCC	780
35	TGGGTGTGAT TGGTTCTTC CTCCTCATTT GGAGCCTTCT CTGCCCTTAC ATTTTGTGTT	840
	CTCCATCTAC CAGCATCCAC CAGTCTATTT ATTAAGTTAG CAAGAGGACA AGTAAAGGGC	900
	CCCTTTGGCT TGAATTTGCT TCTTCTTTTC TGTGGAGGAT ATACTAAGTG CGACTTTGCC	960
40	CTATCCTATT TGGAAATCCC TAACAGAATT GAGTTTCTTA TTAAGGATCC AAAAAGAAAA	1020
	ACAAATGCT AATGAAGCCA TCAGTCAAGG GTCACATGCC AATAACAAT AAATTTTCCA	1080
45	GAGGAATCA ATCCAACTA GACAAATAAA GTAGAGCTTA TGAATGGTT CAGTAAGGAT	1140
	GAGTTGTG TGTTTGTGTT TGTGTGTTT TGTTTTTTA AAGACGGAGT CTCGCTCTGT	1200
	CACTCAGGCT GGATGCACT GGTATGATCT TGGCTCACTG TAAGCTCCGC CTCGCGGCTT	1260
50	CAAGCCATTC TCTGCTTCA GTCTCTGAG TAGCTGGGAT TACAGGTGGG TGCCACCATG	1320
	CCTGGCTAAT TTTTGTGTTT TTAGTAGAGA CAGGGTTTCA CCATGTTGCT CGGGCTGGTC	1380
55	TCAACTCTCT GACCTCTTGA TCCGCTGCT TGGCCTCCC AAATGATGG GATTACAGAT	1440
	GTGAGCCACC CGTGCCTTAG CCAAGGATGA GATTTTAAAT GTATGTTTCA GTTCTGTGTC	1500
	ATGGTTTGGAA GACAGAGTAG GAAGGATATG GAAAAGGTCA TGGGGAAGCA GAGGTGATTC	1560
60	ATGGCTCTGT GAATTTGAGG TGAATGTTTC CTATATGTCT AGGCACTTG TGAAGAATAT	1620

GAGTCAGTTA TTGGCAGCCT TCGAATTATC TTCTCTAGCT TACAATGGAC CTATTGAACT 1680
 GGAAAACACC TTCTTTGAT TCACTTTAAA ATCTCAAAAC TAATTTTAT AATAAAAGTT 1740
 5 TATTTTACCA TTGAAAAAAA AAAAAAATTT AAAAACYCGG GGGGGGCCCC GWACCTTATT 1800
 NGCCCTAAG GGGGGGGTT T 1821

(7) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1024 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGGCAGCT TGAAGAAGCG ACCGAGGAC TUGGAGTGT TAGTGAGGAT GACCCGCGAT 60
 25 GGCAAGAACT GCACCCGAGG GCGGTCTACA CCTACCAGG GAAGAAGAAG GACACAGCG 120
 CCTCGGGCTA TGACACCA; AACATTGAC TGAGCCGGA TCCCGTGAAG GACTTCGACT 180
 GGTGTGTCT CTCCCTGAG CTTGCCAGG ATCTGTGTGT CACCCGAGAT GGCTACCTGT 240
 30 ATGAAGGTGA GGCATCT; GAGTATTC TTAGCAGAA GAAGGAGATT GCGCGGAGA 300
 TAAAGGCTA GGAGAAGCA GGGGACCC GGGGAGGA GAAGAAGSAG CTTCAGCGG 360
 35 CCGCTCGCA GAGCATGT GGGGTTTC TTAGAAGGA GTCCGTATC GTGAGCGGC 420
 CCTCAAGCC TTTCAGAGT AAGGCTCTCT GGGGACCCAG CACAGATGAT GTCCAAGCTG 480
 GGGCAGTGT GGTCTCTCA AGTAAGSACA AGGACAAAGT GTTCCCGAGC TTCTGATCC 540
 40 GTTGGCTGAC GCGGAAGGC AAGGCCACCA AGCTGGAGAA GGTTCGCGC ACGGTGACCT 600
 GGGCATGTG AGGGAAGCT CTGCGATGT GAGCTGAC GCGGTGAGC TTTCAGTGA 660
 45 TAGAAGCTC CTGGAAGC TGTGCTCA TTAGCGGAG GAGGCTAG ATGTCTCTG 720
 TACCTGGA CAGCTTAG AACGAGAG CTGCTCTGT GAGGAGCT GTGAGCTG 780
 TGTACCTCT CGAATGCTG GAGAAGCTGA TCCGAAGGA CATGCTGAC CCTGTGACTG 840
 50 GAGACAACT CACAGAGCGG GACATCATCS TGCTGAGCG GGTGTGACC GTTTCGCGG 900
 CTCCCGAGT AAGCTCAAG GAGAGAAATC ACCGCTGCTG ATGAGGCTT GAGTGTGTG 960

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 983 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CGACACGGGT GCGAGAAGAC GATAGAAGGG CCCGACCGCG AGCCTCCAG GTCTCAGTGC 60
 TGTGCCCGCC CCAGAGCCTA GAGGATGTTT CATGGGATCC CAGCCACGCC GGGCATAGGA 120
 GCCCCCTGGA ACAAGCCGGA GCTGTATGAG GAAGTGAAGT TGTACAAGAA CGCCCGGAG 180
 AGGAGAGAAAT AAGACAACAT GGCAGAGTGG TTTCGGGTGG TGAAGACAAT GCAAGGCTTG 240
 GAGAAGGCTT ACATCAAGGA CTCTCTCTCC CCCAGGAGT AACTGTGAGC CTGCTCCCGG 300
 CTCTCTCTCC AATACAAAGG TGCTCTCAGG CAGGTCCAGG GCTCAGAAAT CAGCTCTATT 360
 GAGCAATTCT GCGGCAAGTT CGGCTGAGC TGCCCGGTGG CATTGAGCG GATCAAGGAG 420
 GACCGGCGCA TCACCATCAA GGAAGACAAG GGCAACCTCA ACCGTGCAAT CGCAGAGCTG 480
 GTCTGCTCTT TCATCAGGCT CATGACACAAG CTGCGCTCTG AGATCCCGGC CATGATGAG 540
 ATCCAGCGCG ACCTGCGAGA GCTGATGAG AGCATGCAAC GCATGAGCCA CCTCCCAACC 600
 GACTTTGAGG GCGGCCAGAC GGTCAAGCAG TGGCTGAGAG CCGTGAAGGG CATGTGGGG 660
 TCAGATGAGC TGGACGATTC ACAGGTGCTT CAGATGCTGT TGGACCTGGA CTCAGCTTAC 720
 AAGGCTTTCA ACCGCTTCTT GCATGCTTCA GCGCGGAGGA CTAGCCCTTG CACAGAAGGG 780
 CAGAGTCTTA GCGGATGCTT CCGTGTCTCC TGTCCGCCAC ACAGGCCGTG GTCATCCACA 840
 CAACTTACTG TGTGAGCTG CCGTCTCTGT GTCTGTCTTT GGTGTGAGAA CTTTGGGGCC 900
 GGGCCCTTCC CCACAATAAA GATGCTCTCC GACCTTCAAA AAAAAAAAAA AAAAAAAGR 960
 KGGGGCGGGT CCCCANTCCC CCC 983

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2421 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCGGCTGATC GCTGCCGCTC CGCCAATACA ATAGAGCCAK CCACTACCAG CAGCCTGGCC 60

	CTCTTCTTCC TTCTCCAGAG AGACCAATCC AGCCGAACTC GGGGTTTGCC TGAGGAGAAG	120
	GAGGAAGTGA CCATGACAC AAGTGAAAC AGACTGAAA ATGATGTTCC AGAAGCTCC	180
5	ATGCGTATTG CAGACCAAGT CAGCAATGAT GACCGCTGGG AGGGCAGTGT TGAAGATGAG	240
	GAGAGAAAG AGAGCTGCTT GCCCAAATCA TTCAAGAGGA AGATCTGCTT TGTCTCAGTT	300
10	ACCAAGGGGG TGGCAGTGG AAACAGTGAC ACAGAGGGGG GGCAGCCTGG TGGGAAAGGA	360
	CGCTGGGGAG CCAGCACAGC CACCACACAG AAGAAAGCTT CCATCAGTAT CAGCACTGAA	420
	TCACTAAAGA GCTTCATCTC CGACATCAAA CCCCTGGGGG GGCAGGAGGG TGTCTGAGAT	480
15	CTTCATGCTG ATACTCTGCT CATCTCTGAG GATGAGACAG ACGTAATGG CCATGATGG	540
	ACCCATGACA AAGGCTGAA AATATGCGGG AAGTCACTC AGGTAGTACC TGCAGAGGGC	600
	CAGGAGAATG GGTAGAGGGA AGAAGAGGAA GAAGAGAGG AACCTGAAG AGAATCTCT	660
20	GTACTCTCCC AGTCTCACT AGAGGTGGCC TTGCCCCCAG CTGCAGAGGA TGAATTAAG	720
	AAAGTGACTT TAGGAGATAC CTTAAGTGA CGTTCCATTA GGCAGGAGAA GTCCGAGTT	780
25	TCCATTACCA TTGATGAGCC AGTCCGAAC CCCAGGTGC CCTCCCGACC CCGGGGCAAG	840
	ATTAGCAACA TTCTCATAT CTCCAATTG GTCCGTCTT TCACTTTAGG CCAGCTAAAG	900
30	GAGTTGTGG GGGCACAGG AACCTTGGTG GAAGAGGCTT TCTGGATTGA TAAGATCAA	960
	TCTCATGCTT TTGTAAGTA CTCAACAGTA GAGGAAGCTG TTGGCACCGG TACAGCTCTG	1020
	CACGGGTCA AATGGCCCCA GTGCAATCCC AAATTCCTTT GTGCTGACTA TGGGAGCAA	1080
35	GATGAGCTGG ATATGAGCG AGGCTCTTTG GTGACCGTC CCTCTGAAAC TAAGACAGAG	1140
	GAGTAGGGA TACATGCTC CTGCAACCC CACCCGAGT CCGCGGTCCA CCGACCAAG	1200
40	CACCCCGGG CAGAGTAGG GAGGAGGAA CGGGCAGTGC GAGAACAGTG TGCAGAACG	1260
	GAACGGGAAA TGGAGCGGG GAGCGGACT CGATCAGAG GTGAATGGGA TGGGAGCAA	1320
	GTGAGAGAAG GGGCTCTTC CGGATCAAGG TCGGCTGAG GCGGCGGAA GGAAGCTGG	1380
45	AAGTATAAG AAAGGAGAG TGAGAGGAAA GAGGAGGAG AGCAGGAGCC AACTGCGAG	1440
	CTGCTGATG AGTTTCTGG AAAGAGGAG GAGCTCTCT GATCTATTG GTTCTCAGT	1500
50	ACTGACAGCC AGATCTTCA GAAAGAGGCA GAGCGGGG AACGGGCGAA GAGGAGGAG	1560
	AAGCGCGAA AGGAGGAGG AGAAGAGAG CAAAGGAGC GAGAGAGGA AGCGAGGAG	1620
	GAACGGAGCC GAGAGTGGG GCGAGAGGA CTTGCGAGC AAGTCTGGA GAGGAGAGG	1680

AGGTACCAGC CACTGGGCCC CAGGGGGTTA TGGCCACAGA GGGATAGGCA CAGTCTCCAC 1920
 CAGCCTGGAG CCAAGGGTCT TDCACATCAG CTATCCCTAC ATACATACCA AATGSAAGG 1930
 5 TGGCCATCCT TTTCGCCCA AACACACCCC CTAAACCTAT CTCCTGGGAC TTAGCCCGAC 2040
 CCTCCCTCTC ATTCCCATTT AASTCTGAGA GGCAAGAGCT AGGTTAGGCA AGGAGGTGCT 2100
 TGGCCAGAGA TGGGAACAG CCAGSTGCTT CAGTCTCTG ATTTTCTCTC CATCCTGCTT 2160
 10 ACCACCTTCC TGGTACTTA CAGCCTTCTC TTGGGAACAG CCGGGGCGAG GACTGGGTCA 2220
 CCTATGAGCT GAATCAGCAT CTCCTCTGTA GTCCAGGGC CCTGAGTTC CCACTCTCT 2280
 15 TCTGTCTGTC AGCCCTTGCC TCTTTCCAC AGGTTCCACT TTATATCCAC CTTTTCTTT 2340
 TGTTCATTT TTATTTTAT TTTTTTATT ATTAAATGAT GTGGTCTATG GAAAAAAAAA 2400
 TAAAAATCTG ACTTAGTTTT A 2421
 20

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 840 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CTCAAACTCC TGAGCTGAAG CGATCTACCT GCCTCAGCTA GGATTACAGG TGTGAGCCAC 50
 35 CGCACCCAAC CTCATAAGC KTATTTGATA AAKATATGC AAGCTCCCTT TATTCACCTT 120
 TCATTCAGAA TGTTTAGTAA TTTTATTGT TTTTCAGATT TTCAGCCCA TATATCTCTT 180
 40 TCCCCACTGT GTCAGTGTAT TCTACTAWA CATCATCAGG TGTTCCTGCT ATTGCTGTG 240
 TGATGGAACA CTGCGGTCA TTTTCTGAA AACTGCCGAT AGTGCATAGA RTGCTGGAT 300
 45 GGAAACCAGA AACTTTGAAT TCAAGCCTTG GTTCTGCCTT GTTTTGTCTT GGTGCTCTT 360
 GAGTCAGCTA CATACCTTTT AAAATCTCAA TTTATTAGAA ATTATTCCAA ATCAAAATCA 420
 AATGAGAAGG TATATACAAA AGTCTTTTAT CCCACAATAA ACTATTCAAG AGAGAGCAAA 480
 50 GGAGAGGACA TTTACTCAAC ACCTCTAAA AGGCAGCCAG TGAAATTAGG CATTTTATTT 540
 AATCTCTCTG GCAACTCTGA GAGTAAAGCA TTATTAATCC CATTTTGGCT GTTTAAAGAA 600
 55 ATTATTTGCA CTAGATTCCA GCTGTAGTTT AGYTTGAGAA AAAAAAATCC TGAGATGTGA 660
 ATTACAGCT TTCTGGGTTT AAAGCCCAAG CTCTATCACA TCATGCTATT ATTGTTACAT 720
 TACTGCTAGT TCTATGAAAA GAAATACTAA TTTATGAAAT ACATCTTATC CAAAAAATA 780
 60 AAAAAAATAC TGGGAGGGGG GGCCCTACC CAAATCCCG GATAGTGATC GTAAACAATC 840

5 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2432 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15 GGCACGAGGC CCGAAGCGT GAGGAAGGGC CCGTCCCGCC TTCCCCGGCG CCCCATGAG 60
 CCCCAGGGCG TTGCAGAGC CCGGAGAGG GGTGAGGAGG ATGTGATTAT GGAAGCTGTG 120
 CCGTATACAA ACTAGAGCA CTCTAGAGC TTCACTTTT ATGATGCCA ACAGGAGGAC 180
 20 CGCAAGAGAC TGGGAGTTG CTGCTCTCGG TCGTGAAGA GGCTTGCCA CCGTCCACCC 240
 GTGTATTTG GGTGCAAGT GTCCAAATC TTCTCCGGA CCGCAACTGC CTGGACCGGT 300
 25 TCACCAAGCG CAGAGCTTG CAGGCATTA GCTGYTATY TGACATCTCT GTCTCTGAGG 360
 GGTCCGTCCC AGATCAGA GATATGGATG TTGTACTGGA GTCCCTAAG TGGCTGTGCA 420
 AACTCGTGGT CAGCAGCGT CTGGACAGA TGTGTCAGC AGAGGCTCGC CTATCGGTGA 480
 30 AGCTCACAGA GCTGTAGGG CTGTACCGTG AGAGAGCTT CCGCAAGAT GTCCAGTTCT 540
 TTGATTGGG GCTCTTTT CTCTAACCG CATTCGAGC CTATCTGCG CCAAGCTCTT 600
 35 TCAGGAGTGG AAAGGATGTC GCTGTCTAAC TGACACATG GAGCTGACCG TGGGGGTGAC 660
 TCGTGAAGGG AAGCCGCCAC CCACGCTCCT TCTTCCCAA GAGACTGAGC GGGCCATGGA 720
 GATCTTCAAA GCTCTCTCA ACATCACCTT GACTCTGAT AAGGGGAGG TGGACGAGGA 780
 40 AGACCGTGGC CTTTACTGAC AACTGGGGAC CTTCTCCCG GACTGTGTGA TATCGCTAC 840
 TCGTGGAGAG CGTACAGAG ATTCCACCG CCAAGGATA AACTCTCTG GGAATTTGG 900
 45 CTTCAAGTGT CTGATGTC TCTCAAGCT GAGCGATAT GAGACTGCA CGGATTCAT 960
 GGGAGTGAAT ATGATGTA TTGCTGGCT CAGATCTTC CTAGAGAAC GTTTCACAA 1020
 GACACACAGG CTGAACGAGA GTGTAGCTCC CCGTGTGAGC CTGCTGACAG AATGTGCCCC 1080
 50 GATGCACCGC CCAGCCAGGA AGTTCTTGAA GAGCCAGTG CTGCTCCCTC TGGGGATGT 1140
 CAGGACACCG CTTGAGTTG GCGAGATGCT GCGAGAACG CTTGTGCGC TATGAGACA 1200

60

ATGAGTACAA GGAAGCCAAA GCCAGCATAA ACCCTGTGAC CGGGAGGXTG GAGGAGAAGC 1440
 CGCCTAACCC TATGAGGGG ATGACAGAGG ASCAGAAGGA GCACGAGGCC ATGAGGCTGG 1500
 5 TGACCATGTT TGACAAGGTC TCCAGGAACA GAGTCATCCA GCCAATGGGG ATGAGTCCCC 1560
 GGGGTTCATCT TACGTCTCTG CAGGATGCCA TGTGGGAGAC TATGAGGAG CAGCTCTCCT 1620
 CGGACCCCTGA CTCGGACCTT GACTGAGGAT GGCAGCTCTT CTGCTCCGCC ATCAGGACTG 1680
 10 GTGTGTGCTT CAGAGACTTC CTTGGGGTGG CAACCTGGGG AAGGCACATC CCACTGGATC 1740
 GACACCCGCC CCCACTTCTC CATCTTAGAA ACCCCTTCTC TTGACTGCCG TTGTGTTTAT 1800
 15 GATTTGCTTC TGGTCCAGTT TCTCATCTCT GGACTGCAAC GGTCTTCTTG TGTAGAACT 1860
 CAGGCTCAGC CTCGAATTCC AAGACGAAG TACTTTCTTT TGCTGGGCC AAGAGGAATG 1920
 TGTTCAGAAG CTGCTGCTTG AGGGCAGGGC CTACCTGGGC ACACAGAAGA GCATATGGGA 1980
 20 GGGCAGGGGT TTGGGTGTGG GTGCACACAA AGCAAGCACC ATCTGGGATT GGCACACTGG 2040
 CAGAGCMANT GTKTTGGGGT ATGTGCTCCA CTTCGCCAGG AAAAACTG TTAGAACTTT 2100
 25 CCATACGAGT ATATCAGAAG ACACCTTCC AAGGTATGTA TGTCTGTTG TTCTGTCTT 2160
 GTCTTCACTG AGCCGAGGGC TGGAGGCTC TTAGACATTC TCTTGCTCC TCGTTCAGCT 2220
 GCCCACTGTA STATCCACAG TGCCCGAGTT CTGCTGCTT TTGGCAATTA AACCTCCTTC 2280
 30 CTACTGGTTT AGACTACACT TACAACAAGG AAAATGCCCC TCGTGTGACC ATAGATTGAG 2340
 ATTTATACCA CATACACAC ATAGCCACAG AACATCATC TTGAAATATA GAAGAGTTTT 2400
 35 GGACAAAAAA AAAAAAAAAA AAAAAAAAAA AA 2432

40 (2) INFORMATION FOR SEQ ID NO: 49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1742 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50 GTCTGCAGG AGCTGCACGC GCCCGAGGTG CGCANGAACA AGGAGCAGCG AGAAGAGATG 60
 TCGGGCTAAG GGGCGGSAC GGGGGGGGCT CATCCTGCCA CGGAACACGT TCGGGTTTTG 120
 GTTTTGTTTT GTTCACCTCT GTCTAGATGC AACTTTTGTG CTCTCTCCGC CACCCAGCC 180
 55 CCCAGCTTCA TGCTTCTCTT CCGCACTCAG CCGCCCTGCC CTCTCTCTGT GGTGAGTCGC 240
 TGACCACGGC TTCCCTTGCA GGAGCCGCCG GGGGTGAGA CGGGTCTCT CCGTGCAGAC 300
 60 ACCAGGCCGG GGGGGGCTGG GTCCCGGGG GGGCTGTGA GAGAGGTGGY GGTGACCGTG 360

	GTAAACCCAG GCGGCTGCG TGGGATCG GGTCTTACG CTGGGCTGTC TGCTCAGGAC	420
5	GTGAGGTTGA GGGGAGGTTC TGTGAGCGCG GCGGCTGCG CAGGAGGCGA GGCTACAGTA	480
	CTGTCTGTCT TTGAGGCGG AAGGGCTTC CATTGAGGTA GGGGCGAGCG GGGAGGTTGG	540
	TGATGTTGCG TGGGAGGCT GGTGTGCGAN GGGGTGCTTG TTGAAGTGGC AGGCGGTTGG	600
10	GTGGGGGTTG CAGCTTTTCT TAATGTGCTT GCACAGGCTT CTTCTAGAC CAGCTGCGCT	660
	GAGGTGAGCA CCGTGGGCTT TCGTGAAGC CTGCACTGCG GGGCTGCGC TGAGTGTCTT	720
15	GGGGAGTTGG CATTCTCTGC CAGGGAGCGA TGAGGAGCTT GATGCTGTA GAGGTTGTGG	780
	GCAGGATGGA CAGTGGGCGA CTCAGAGGTC CAAGAGTTC AAAGAGGCTC TGCGCGAGG	840
	CGCTCTGTGG GAGAGGCGG GCGGCTGCG CAGCAGGCG TTTCAGATG TCGTTGAAAG	900
20	AGCGAGGCTA GAGGCTTTTG GAGTCTGCG GCGGCTGCG CCGTCTGCG TGCTGGAAG	960
	GGCAGCAGAA GTCTCTGTA GGGAGCGCGA AGGGGATTT TGTGGAGCG CTGCGGAGAG	1020
25	ATCGAGGTTT TGGAGGCGA GTGGGTAAGG TTGCGAGG AGCGCGAACA CCGTTGCGAC	1080
	TTGGCAGCGA GAGGGGCGTG TGGGTGAGG CTTGACTGCA GGGCTCTCTT GCGCAGCGC	1140
	TCTGGGCTGA GTCTCTTCTT TCGTTGAGC GCGCAGGCTT GCGCTTGGAG GAGGCTCAG	1200
30	TGGAGGATGG GGGTGGGCGA GCGTGTCTTT GTACCACTGC AGCATCGCGC ACTTCTGAC	1260
	GGAAGCGCGA TCGCAAGGT GCTGCTGCG CCGTTGCTGT AAGTGTGAA GGGGGGCTT	1320
35	GAGTCTCTTT AGGAGCGCA GCGAGGCGCG TCAATTGCA TCGTGGGCGA GCGCTTGGCG	1380
	GGGCACTGCG ACTGTCTTGC AGAGCGCAGC CAGGGAGTA GGGGAGGATC CTGACCGCTG	1440
	CAGGGCTGAG GGTGAGGAG GAGCGCACTG CCGCATCTGC CTCTCGCGAC CAAGACAGCT	1500
40	CCAGAAGGAG CAGGAGGTC GATGGGAAAC CCAAGGCTGT CCACATCTGG CTTTTGTGG	1560
	ACTCAGAAAG GGAAGGAGAA CTGAGGCTG GATATTTCT CATGCTGCGA GCGCTCATAG	1620
45	CGAAAGGCTA CTGTAATAG CAGGATCTGC ATCCAGTAG TAACTGAAC TTAAAAATG	1680
	AATCAAAATGA AATTAATTAAT AAGACCTGT GTCTTTAAGA AAAAAAAAAA AAAAAAACTG	1740
50	CG	1742

(2) INFORMATION FOR SEQ ID NO: 50:

SEQUENCE: 1742 bp
 (3) TECHNOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

5 GGCACGAGCC TCGGTGAACCT GTGGAGTCGG GGGAGGGCTG GAATCAGCGT GGGCTCCAGG 60
 TOGCTGGCAG CCGGCTGACA GAACTCTTCC GAGGCTCCTT GGGGAAGAAGC TACACCCGAG 120
 GAGGCTGGAT GGGGCTGAAA AACCTGGCCC GCTCTGGTTC TGTACCATTC CAAGGGGAAC 180
 CGTAAACTGA GCTTTTCTAA CGTGGGTTTC TGCAAGTAC TTTTCAGCT GCCCCTTCC 240
 CCCCAGACCA CAGGAGAGCC TCTGTGTAGC GAGGCTTGA CAGTGGTTAG GTAGGTTSTA 300
 CTGTGTAGGG AGGAGCTCAA GATCATGAAT GGTGTGCACA GAGAAAAGCG GTTGCTCTTT 360
 15 TGCAAAACTA TATACCTGCT GTGGTTTGTG TTTCTTTTC TCTGAGTAA TGAAGTTSTA 420
 AGTTCAACTT GGCACATTTT CAGGCTCTG CAGATTATTT GCACTTTATT TCATAGGTCR 480
 ATAAGTACTT TTAGCTTTTC TTTGTATATT GAGTGGCTTT TGAATTGCTT CCGATATTTT 540
 20 TATTTCAATC AACTGAACA ATTGTGGGCC CTCTATTTTA TTTATAAAGG TTCAGTSTAT 600
 CTTTGGCTTC CTATATCAAT CTGCAAGGGA GTTCAGAAA GCTCTATGTT CATCGAGCCG 660
 25 TGAGTCACAA CCAATTTTGA AGCTTTTATA AAAAAAAGT GTTTGCTTTT TTTCACAAGT 720
 AACTTTAAAA GTGTAGTTTA GAAAGAAAAC ATTTTCAATA AAAAGACACT ACATTAATCC 780
 30 TGGATGCTTG CAAATCTTAA AATMTATTCG TCTCTAGCG TTGCACAGCT CTGTGTTSTA 840
 TACACAGACT AGCTTTAAAA TTTGTCAAT ACCACTTTAC CTTTACTTTT ATGTATCATT 900
 CCCCAGACTT CTTACTGCA GGTGTGGGCA AGAAAACTTT TCTTTAACA CTTTCAACA 960
 35 GGGGGCATAA AATCTGAG CTGAGGTCTT GAAGAATGCA GATGGGTACA GTATGTGTTG 1020
 GAGCTCACAG TGTGTAITGA CTAACCTAGT TCCTTTTTTG CTTTTTTTTG TATTGCTTG 1080
 40 TTAAAGTGA CTECCAGSTA GCAACTCTCT TTTTAAGGG TGGGAACGAA AGGACGTAG 1140
 GAAGAATAGA TCTAGATTAT TTAACAGTCT TCGATAGAGT TTGAAAGCTT TCTTCTTCAT 1200
 TCAATTTTGG GCAAAATACT GCCTCTGCAT TTGTTCATAA CAAAAAGATT AGATTAATAA 1260
 45 GTAGCTTTTG TTGGTGGAAA TTACCAGCTC TATAAGTCAC CCTTGGTGGT TCATGSACCT 1320
 CTGATTAGCT TGGGTTTTGC AGTCTCATTC CCACATGTAT ATGTGGAGCC AATGGCTTTT 1380
 50 TGGTGCTCAG CTGTTTACGT CTGACTCCTT GACTTCTTTG GTACAGTCAT GGAGTCAGAT 1440
 CTCATTAAAGT GTGATTCTCC ATGGATATAA CCAGCCCCAA AAAAANG 1487

55

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1328 base pairs
 (B) TYPE: nucleic acid

311

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5 GGCACGAGCT GGTGGCGAAT TGGGACGAG AGAAGATTTS AAGAAGCCAG ATCCAGCTTC 60
 CCTGGGGGCT GCTTCTTGTG GGGAAGGGAA AAAGAGGAA GCTGTGAAGA ACTGCACCTG 120
 10 TGGGCTTGCC GAAGAACTGG AAAAAAGAA GTCAAGGGAA CAGATGAGCT CCGAAGCCAA 180
 GTCAGCTTGT GGAACCTGCT ACGTGGGCGA TGGCTTCCGC TGTGCCAGCT GGGCTACCT 240
 TGGGATGCGA GCTTTCAAAC CTGGGAAAA GGTGCTTCTG ATGATAGCA ATCTTCATGA 300
 15 TGCTTAGGAG GTTCTTGACA TGGGACCCAT CTGCTCTCC AGCCAACTCC TGTGCTTCAC 360
 ATCCACCCAT GGTGGCTGCT CCGACCTGCT CTGATTTGT TCACTCTGAG ATCTGTTTGC 420
 20 AGATGGGCTG CTTAGCAGAC AGAGTGAAGC TGGCTGGGGG GAGAGTGGT GTGTAGTGGT 480
 GCTGTGTATC AAAAGACCAA GGTATTATGG GACCTGGTTT CAGAATGGGA TGGGTTTCTT 540
 CACCTCATGT TAAGAGAAGG GAGTGTGTTC TGAAGAAGCC CTCTTTCTGA TGTTAAAAAG 600
 25 CTGACCAGAA GCTCTTTGAG CCGAGGATC GTTGAGGATT AACACTCTGT GACAGAGCTG 660
 CAGAGGCTG CATTGAGTGT CATCTCAGCA ATGCTGCCAC CTCTTTGTGT TPCAGAGTTG 720
 30 TTAGTTFACT CATTCTTTG TACACGAGT CAAGTGGTTC ACAAAGCTCT CAGGGCACCA 780
 GAGGACTCAC TCACTGGTGG CTGTGATGAT ATCCAGTGTG CTCTGGGCCC CTTCATGCCC 840
 CAACACATTT TCACTGTAGC ATGCACTGTG TGTCTGTGTT TCAATTATGT TAAGCTTCAG 900
 35 GTATTAAAST TGCTGCATAT CTTCACATAT CTGAGATTG TGCATGTCTT GTAAAGAGAG 960
 GGGATGTCCA TTGCTGTGTG ATGTTGGATA GTCATCCAGC CTGAGTTTGG ACCATTGAG 1020
 40 GAACTTAGTG TCAAGACAAA ATGGGGCTAT TCTACGCTT AGAATAGGGC TTGTCTGGCC 1080
 ACTTTAGAAG ATCTCCAGCT TGGTGAACAT TTAGAGGAA GCAGGGCAGA ACTCTGAAG 1140
 AGAATACCTC TCTCTGAGCA GAGACCCCTT TGTCTTCTT ATCCACCCAT AAGGACTGG 1200
 45 AATCAAGCTT GCGAAATATT TGGAGAGATT GTGTGGATTT AAGAGACCTG GATTTTATA 1260
 TTTTACAGT AATAAAAGT TTTCAATGAT ATCTGTCTT GAAAAAAAAA AAAAAAAAAA 1320
 50 AAATCGA 1328

60
 (E) TYPE: nucleic acid
 (F) STRANDEDNESS: double

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	GAATTGGGCA CGAGCTTTGC AACATGCGAA ACGACTTGGT AATCGAGGGT TCCGCTGCCC	60
	CCTAGATTAA ATTCGCGGGG CTGAATTTGA TTGCGAATT TACAATATCA TATTTTAAAT	120
	TGCTGTCTTC AATTAAAGCA TTATGAGCA TAACTAATT TCGGATGTC GATGCTAGCT	180
10	TTTCAGGGG TTCTTTTCTT GTACAAAT AATGTGAT AAGCGTTTC ACTTATATTC	240
	TTCAACATG ATGCTAATTT AATTAAITA CTTCCTATGA TATTTATCA TTCTATGAT	300
15	TTTGCCACTG TTATTATTC TCTCAAAAT ACATCTAGG AAGAGGATTA TTMTAAGTGA	360
	TTTGATTATC TTCTATCTC TTATTATTAT TTCTGATTA CTAAAGAAAT TCGTTGCAAT	420
	GGTTGCAATT GATACAGTAA ATTTGTAAAT GAGGAGACAA TATAAAAAAT CTAAATTACT	480
20	TTTGTCTAAT GACTGTAGCA GAATSCCTTT TCTCTAATC AATTTGCTTT TCTTGCAGTT	540
	TAGTTTGATA GATTTGCAAG CTATGCTGCT TCCATGAAT TACTGCGCT GGTAGCAACG	600
25	CAGGCTTCTT TTTCTCTGCT TGTAGCTTGC ATGATCGGCT TATTAGGCA ACAAAGTAGC	660
	CGGAGATCAC AATTCAGGTC CTTCGGTATG TTGCTATCT TTGAGGTTGC AGAGAGGTTG	720
	GGAGAACTG AACTCACTGG CGAAGCTGG CATTGCACT GATTCTTTAA TGCATCTAT	780
30	GTCTTCAGGA AGCCACAGGC CATATTTGAC TTGCAATAG AAAAAGAGAG GAAAAACCCC	840
	ACAAAGTATA AGAACCCCTT AAGATATATC TATTTTAAAG TGAATTAAT TTMTGAGTTT	900
35	ATACCAATGG CGAATTACAA TATAAAAAAT TTCAATTTCT TGAAGATCC TTTGTTGACT	960
	TGTCTTTTCA TCTCTGCTA TTATATTTTG TCACTGTTAG TCAACAACT CTTATTTGCT	1020
	GAGGAAGGAC TTGCTGTGAC TTACTGTAGC ACATCAACA CTGGGAGGG TGGTSTTTAA	1080
40	CTTTTAAAAA AATGTTATTC TGATTATAG AATATATTC GCTTTTTC A TGAAGAGAGC	1140
	GCCACCTTGC AAGCTTTAGT GAGATTTATG GAGTTGAAI AACTAGGAG GAATTGCTGC	1200
45	TAGTTCACAA AATTTGCGAA GCAAAAGCTA GCGCCAAITG GTTGGAAGT TTGAAACTGA	1260
	TTAACAGATT TGCATTTGAA CTGACTTCAG ACATTAGGTC CAGACATTAG TTAAAAATAG	1320
	AAAGAGGAAT AAAGACATCT YTTCTCTCTA GAAAGATAA GACTGCAAT AATAATCCTT	1380
50	CCCACCTTCA TTGAGATCAG CTTCCTGAT AACCTGATAT GATGTTGATA ATGATAAACA	1440
	TGATAATAGT GGTACTTTTG TAATTTTGCT GGTGCAATTA AAGAGATACT AAAGATGAG	1500
55	TTCAVCTTTT CTTCGAACAT YCTATTCCT AGATGTAGT TACCTCAAT TGGGAATTAT	1560
	AACTGTCTTA ATTTTGTG TGACCTGGA TCGCCCTTCT GCTTAATAC CCACAGTGTA	1620
60	ACAATTAAAT ATCACACTAT GACATATGAT TTAAGTAGCA TATTTTAAAG ATAAATTTTA	1680

GGGGTAAATG TTTACTTCAA AATGACTCCA TATTTCAAAT ATCTTTTAA ACTGTGAAG 1740
 CCAATAAAT TTTAAGAAAA CATTGAAGA GTAGTGTGTT TGCATTTGT AATAATCTTA 1800
 5 CTCACAGCAA GTAAACGTAA TAAAGCCAA CATTTAAGCC AAAAAAAAAA AAAAAA 1856

10 (i) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1558 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

20 TGGGTATCCA TTCTGHAAT TACTTTACTT AGGATAATGG CCTCCAGCTC CGTCCAAGTT 60
 CCTGCAAAAG GTATTATTTC GTTCTTTTTT GTGCTGAGT AGTATCCAT GGTGTATATA 120
 TACCACATTT TCTTTATCCA CTCATTGCTT GATGGGCACT TAGGTGGTT CCACATCTTT 180
 25 GCAATTGTGA TTGTGCTGC TCCAGATATC ATCTTTAACT CTTTGGCTT CTCCACATAC 240
 ATTCCAAGT CCGTTCATT CTACTTCAA AATSTATCTT GTATCCATC ATCTCTCTCT 300
 30 ATCTTCAATC TATTTCAATG CCGATCATC TCTTCATCG AGGACTGTAA TAATTGGCTA 360
 ACTGGCTCTT TCTTACATTT TAAATCAAA AGATGTGACA GGTGAAATGC CTATTTCACT 420
 GTCCATTGAT GGTCTCTCTT ACAATACACC TGGCTGGCTG GTGTGGCACT GGCAGAGTTC 480
 35 AGCAGTGTGA AAAAGACTGC TTGCTCTTTT ACAGGGAAG CAGTTCAGT GTGCTCTGTG 540
 AGGACAGAG CTCTGGCAG GTCTGGACAC TGGCAGACC TGGTCTTGG TGGCCAAGGT 600
 40 AGCAGGTAT GTTTTGGG TACTCACAG GCTCAGAC CACTCTCAT GGTCTCTTA 660
 CTCTTCTGCG AGAGCTGAG CCGCGGTGA TTGATCTCT CTCCATATG CTGTCCATG 720
 GTTCTCTGA TTAAGGGGCT TCTTCACAA GCTCTCTGA GACCAAGAG TATGACATG 780
 45 GAGCGCTCT GACAGCATC CACTATTCAA AGCATGCTG GCTCTCTGA CCACTTTTCT 840
 CCACCTCTTC TGGTGGCTG TCTCTGTCT CATATCTGT TTAAGCTTC GTAGAATTGC 900
 50 AGGTCTCTGT ACGGGCCAGT TTCTCTGCT TCTTCAGGA TCAGGGGTTA GGTGCAAGA 960
 AGCATTTAG GCGACAAAA CAGTGCAT GACGAGAGG TCTGTCTG TCTCTCTGCT 1020

1060 TCTATTAATTT TCTCTATATG AATGATATA TCTATTTT AATTAATA AATTAATTT 1120

TTGTCAGCAG GCAGGCTGGG GAGGCCAGTG TTGTGGGCTT CTTGCTGGCA CTGAGAAGGC 1320
 TCACGAAGGG CATCGCAAT GTTGTATTCA CTGAGAGCTG CTTCTGCTC TCTTCACCAC 1380
 5 TGTAGTTCTC TCAATTCGAA ACCATCAGCT GCTTTTAAAA TAAGATCTCT TTGTAGGCAT 1440
 CTTGTAAAT TTGTAAACAA TTAATTTAAA TGGCATCAGC ACTTTAACCA AAAAAAAAAA 1500
 10 AAAAAAAAAA AAANAAAAAA AAAAGGGGGC CCGCTTAGAG GTCCAASTTA NGACGNGG 1558

15 (2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 948 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

25 TAAAAATCAT GCTGTGTAGC ATCTCAGCG TAGTCATCAT CATGCGCGGG CAGACCAACA 60
 GAACTACTGG GATGCTATAA AAGGCGCTTG GTGCGCGCGC ATTCTGCGGC CCTGATCTC 120
 30 CAGGCTCTT TGTGAGWCA TACGCGGAC CCAATGCGG CCTGAGACAC CCGTTCTTG 180
 GCGCGTACA CTGCGATACA TGTAAACTC CGCTTCAAG GAAGTCTTG CTTGCGAGC 240
 AAGTTCGAA TCGATTCTC CAGGAAGCG TCGAAAAGC AGACCGCAG GGAAGCGCT 300
 35 TTGCGGATC CGGCGCAAC GCGGAGCTT CAGTGGGTC AGGCGCGCT ACCGTCAAAG 360
 TGTAGCGGC CCAACGAGC AAGCTGGTT TGTGCTAA AACCGCGCT CTTTATAG 420
 40 CACGCGCGCA GCTCTGACAA AACCGCGCT CAGGTGGG AGGCTCGCT TCTTTCTTC 480
 TCGCGGGGT GATTCACTC AGTGATTGG TTGTGGCTC CAGGCTCGC CCACAGACG 540
 ACAGACCGT CCGTTCTTC CGCAAAAAG ACGAGCGCT GGGGTAGTAA GGGCGGACA 600
 45 CTCTGTTTT TTGCAAGTAC ATTTTGTTC YCTTCCAC CAGGTACTG CTTATTTCT 660
 TCTAATGCG AGAAGCTTC CTTTGTCTT TTTAAGGAC ATTTGGGAG TTCCTGCTT 720
 50 AGGACCGTTC TCGTGGAT AAGAAAGTG CTTGTAAAG CTCTGTAAAT ACTCGCTTC 780
 ACGCATCGA GCGCGTGGC AGCGGGGAG AAGGGAATC AGGCTATGA CTTCCCAAT 840
 CCGCGCTCG CCGTCCCTC GCGGGCGCG CTTGTCTG ATCTGTGTGT GAGTGTGTGT 900
 55 GAACTCTGA AAGACAATAT TAAAGAGCT TAGTTGAAAA AAAAAAAA 948

60 (2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 990 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10 GGGGAAGTCC AGTCACAGCA GSAGTANGAG TGGGAGGCAG GACAGAGCTG GGACACAGGT 60
ATAGAGAGGG GGTTCAGCGA CCTAGAGAG GGCAGACTAT CAGGGTGTCC GCGGTGAGAA 120
TCCAGGGAGA GSAGCGGAAA CAGAGAGGGG GTAGAAGACC GGGGCACTTG TGGGTTGAG 180
15 AGCCCCTCAG CCAATTTGGG AGAGAGGCA CATTGGCTAC CAGGTCCCTT ACACATCCC 240
GGGTGCGCTT TGGTTCTGGT GCTTCTGGG CTGGGCGCGG GGTGGGCGCA GGAGGGGTCA 300
20 GAGCGGTGTC TGTGAGAGG GAGTGGCT GGGTTGTG AGCTGGCG AGCTGGTCA 360
GGGGGCGCGG GGGGAGCAGG CTGGGAGAG GAGCCCCCTG GGAGAGTGGG ATTTGTTGG 420
GTGGAAGGC AATACCATTA GGCAGCAGG GAAACCGGCA ATGGCACCAG TGGGCGCAT 480
25 TACTTGGAGC AGTCTGTGT GAAGSAGGG GTTGGCTTTG ACCGGGCGCTG TGCTCCTTC 540
GTAGCCCCCTG TCGGGGTGT CTACAGCTTC CGTTCCATG TGTGAAGGT GTACAACCG 600
30 CAAACTGTTC AGGTGAGCT GATCTTGAAG AGGTGGCTG TCATCTCAGC CTTTCCCAAT 660
GATCCTGAGG TGACCGGGA GACAGCAGG AGTCTGTGC TACTGCCCTT GAGCCCTGGG 720
GACCGAGTGT CTCTCGCTT GGTGGGCGG NAATCTACTG GGTGGTTGGA AATACTCAAG 780
35 TTTCTCTGGC TTCTCATAT TCTCTCTCTG AAGGACCCAA GTTTTCAAG CACAAGAATC 840
CAGCCCCCTG CAACTTTCTT CTGCTCTCTG TTGCCCCANA AACAGCANAA GCAGANANA 900
40 NACTCCCTCT GGTCTCTATC CCACCTCTTT GCATGGGAAC CTGTGCCAAA CACCCAAGTT 960
TAAGAAAAAA ATAAACTCT GCGATCTCCA 990

45

(i) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

60

CCGCCGCTCCT AGCCGCTGCT GTTTCGTGG GAGCCGCGCT GAGTTGCGG CTGTTGGCTC 180
 CGGACAATGG GAGCAGCGCG AGATTGCACT CCAGAACAGA GAGGACCCCG TCGCCACGCA 240
 5 ACCGATACTGG GAATGGACAC CCAGAATATA TTGCTACCG GCTTGTCCCT GTGTTCTTTA 300
 TCATGGGTCT GTTGGGCTC CTCATTTCG GATTTTCTT NAAGAAGAAA CGGTATCGTT 360
 10 GTACAACAGA AGCAGAGTAA GATATCGAAG AAGAAAAAGG TTGAAAAGWT AGFATTGAAT 420
 GACAGTGTGA ATGAAAAGAG TGACATGTT GGGCAATCG TCCACTACAT CATGAAAAAT 480
 GAAGCGAATG CTGATGTTTT AAAGGGGATG GTACACATA AAGAGTSTA TGATCTGAA 540
 15 AGCCCTGTGA CCCCAGTAC ACCAGGGAGC CCCCAGTGA GTCTDGGCT TTGTCAGAG 600
 GGGGAGCGC AGGGAAGCAC GTCTGTGCG ATATATGTCA TACGTTGGC TGTSTWGTG 660
 AGAGGGATGT GTGTCTCGG TGTAGGACA AGCGTGGCA CTTTATAAAG CCGACTAACA 720
 20 AGTCCAGAGA GAGCAGATCA CGCCGCGAAG GCGAGGTGAG GTCTCTTCT GTGGCAGAT 780
 TTAGAGTNA AAAAGTGAAG CACAAATCAA ACCAGAAGGA ACGGAGAAGC CTGATGTG 840
 25 TTAGTGGGG TGAACCTCT AATGGGGAGG TCGGGCAAC AATGTGAAG AGAGAAGCA 900
 GTGGCAGAGA GTAGCAGTG AATGTGTT TTGTTGACAT TGGGGCAGA GTGTTGCAAG 960
 GTGAGGAGAA GGTACTTGA GCTCCGAGG TGCTGTGGCA GCATAGGAAT GGTATTTGAC 1020
 30 AGGGAAGTGG GAGAGCTTC CTGACCCAG GAAGACTGAG GGGGAGTGA CATGATTACT 1080
 TGTCTGCCA GAGCTTCTG TAAAGAAGTC ACAAACTTAG TSCCTGAGG GGCCTGGCTG 1140
 35 TGTGATAATG AGGATAGAG ATTACTTGT AGGCAATGT CCATGTTGG GATTGTGCA 1200
 AACTAGAATT CACATCACCC AATATATAG GCTTGCATTA CCACGAGGCA GAAAGCACT 1260
 40 AGTGTGCTG CATCTCTTA CCAAAAAAG ACAAAATCCA GACTTCTAAA ATGTAAAATC 1320
 ACTGATTTTC GATATTGCA GCTTACTTT TTTTTTTAAA CAACCATGCA GGCCTAATGA 1380
 CTTGTAATCT TGTCAACATT TTAGGTAAA CTGTGACTTG AAAAAGTCT GAGCAACAA 1440
 45 ACCAATGCTT TTCTCTTTA TTCTGTGGR AACCACTTT CTTTGTGTCA CAGTTTGA 1500
 ACCTCAATAC GAATATTCT CTCCACCA AATATTTGA GGCAATTGAA AAGCCACAGT 1560
 50 GATTATTTT TCGATTGGC AATTTTAATT TTGCAAGACA ATT 1603

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1052 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

5 TACAGGTCAG GATGGCTGTA ACATTGTCAT CTCTGGGCTT CTGGGTGCTG CTTAGGCTGC 60
 TTTTGGCTG GAGGACTGAG GAGGGATGCG GGGGAGCAAC ATGTTACTAA ATCATACTTT 120
 GCTGGCTAGG TTGCGCAGAG CTCTGACTGC TGCGTGGTGT TCGAAGCGGT TGTGTGGGCA 180
 10 GAGTATACAT TTTGGAACCT CTGCGAGGCG AAGCTGCACT TCGAGATGAA CCATAGGCTG 240
 CTCTAGCAGN AAGGCGCGAG ACATGATGCG AGAGGAGCGG AAGAGGAGCG AGCTGGAGAG 300
 GGACGAGGCT ACASTGAGAG AGCAGCTGCT GCGAGAGGCG CTGCAAGTCA CTGGGGAAGC 360
 15 TAGGCTGCGA AGGACAGGCT TGACAAAAT CTGGGCGAGA CTGGAAGAGC GAGTCCAAAG 420
 CTGCTGCGAG GCTTTGGAAC TCAAGCGAGC TGACTGGCTG GCGGCTGTG GCACTGATC 480
 20 AGCTGAATG AGGCTGGGCA CTGCGCACTT TGCGCTGCGC TGTGGCTGCA GCGCTGCTGT 540
 MYGCTGCTT TTCTGCTGA AAGGCACTGC CTCTGCTGAT AATGAATGCT GTTGGCTG 600
 25 CTGCGCTGCG GAGGCGCGCA GCGAGGCTT GCTGCGGATA GATACCTTTG GCGTGGCTG 660
 GACAGGCTGC TGAGGAGGAT TGAGGCTGAA AGTCTGCGAC GAGTACACTA AACCTAGGTC 720
 TGGTACGAA TAGGCTTTG AGAGCAAGG GCGACAATC ATCAGCTGCG TGTCTCTTAG 780
 30 ATGCACTTTT TTTTGCAGC AGTACATGCT TCAACACACA GAATTTGAGG GAAGAGTTCT 840
 GCGCAAAAGC CTAGCTGCTT AGCTTCTGAT TTTAGCTTTC CAGCGAGCTT CCACAAAAGA 900
 TTTGGCTCTA CTTTGGATCT GGTAAAT AACTAATAGG CAGGCACTTA TTTGGTAA 960
 35 GAAAAAAGG CTGGGAGAGA CAGAAAATTT GCGCACTGCT GCTGCTGCGC TTGCTGTG 1020
 AGCTGGATTT TGCTATGAA TCTGTAGCTT NN 1052

(2) INFORMATION FOR SEQ ID NO: 58:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 814 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

ACNCGNTGCG GCGGCTGTA GAAGTAGGCG AAGCGCGCG CTGCAAGGAT TCGGAGGAG 60
 60 TTTTCTTCA GAGTTGCTT TTTTCTTCA GAGTTGCTT TTTTCTTCA GAGTTGCTT 140

ACTGCAACCT GGGAGGCAGA GGTTCAGTG AGTCGAGATG GTCCCATGTC TCTCGTTTGG 300
 GCAACAAGAG TGAAACTCTT GTCTCAAAAA AAAAAAAAAA ATGAGGTTTA AGACAGTTTT 350
 5 GTCATTACTG GTGGGATCTG GTACACAAG ATAGCATTAA ACGTACATG GCACATAAAA 420
 TTGGTTAAAA AAFTTTGTTC TTAAATTACG TAATGTAAAA GCCCAACAAA CACTTTATGC 480
 AAGATTGGAA TGTATCTTCA AATTGAGATT TAATAACAT GTAAAGATCC TGTGTATATA 540
 10 AAAGTTGTAT TTAATCCCTT GTCCCCAAG AATGCTATAA AAGATCCAA GAATGTTATC 600
 TATGAAAAGA TAGCAATAGG GAATGGTGAA CAAATAATTT AATTGCCAA TTCTAAAAAA 660
 15 CATGGACTTA AATCCCATGA AACTTGGTT CCATAGTTTT AACTGTTTTA TGGTTCCAAT 720
 ACAAACCAG AGTGGTTTAC ATTCCACAAT NACCAAAATT GCATCCAAATN TTCCGGTAAT 780
 20 TTTCGGTATT TGCCATGGGA TACTATTCAT TTTT 814

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AGAGGAAGTC TTTCSCAAG CTTGTCTCTT GGACTAACGC CATCCAGGCT GGGAGGGGAA 90
 35 GAGTGCTCTG CTACTCTCTT CCCCCTCTG CTTCTCTTC CTTCTCAGCC TTGGTTCCTG 120
 ATGGGAACAG AATGAGGCTC CTGAGAACAT ACTTCTTAAA TGCCCTTGAC CCAGGAACCG 180
 40 ATTATCTATA TTGTTCCTCA TTTCCTTCA CCGTGACATT CCAGCATTGT CTGACTGTGA 240
 GGTGGGCTTT TGAGAGCCTC CAGGTTCTTC AAAACAGGCC TGAGCGATGG GCATCACACT 300
 45 CTCTGCCTAC CCACTGCTT GTTACCTGC CAGATAACCA AGTGNAGATG TGTGAGATG 360
 GCTAGTTTTC ACATTCTTAC TAGTGTGTTG YTCACCTTTG GGCAGAGGCC CCTCTAGGC 420
 CTTGCCCCAC CTCATCAAA CCGAGACACT GTAGTCAGAC CTCAGYAATA TAGGAGGCAA 480
 50 TAATCTTTTA ACAGTGTTTT GCAACAAAC AAAAGAGAA AAATCCAGC CAGGGAAGT 540
 CGCCACCTGC CAGGCTAGT TCCATCCAGC CTCAAGACC GGCCTTAGAC CAGGAGGGA 600
 AAGGCCCCCA TCACACTCGG CCACTAGTGG GTCCTTGAGG CCAAGAAAGA AACCAGACTC 660
 55 TGTATGACAA GTTGGGTCTT TTCCAGAACA CGACAGAAAC AGGGGGGGCC CTTGTTAAT 720
 GCCACTCCAT ACTCCAGAAG CATTATTCCT TATTGTTGGC AGCCAAGGGC AGATTACAG 780
 60 GTTATTGTAG GAATAAAGAC TAGTTTACAA AGGARAAACA GSCCCTGGAC TTCCCMAGGA 840

AAGGTGAGGT TAGGGCTCCT GTACCCATTC TGTTCACCA CTGTTTGATC TCTCTGGCCT 900
 CCCACGAGGA ATGCCGTTTC CTTTATTATG ATCTGTTGGG AACCAGAGAG AATCAACAGA 960
 TCAATGACAT AGGATCCGAA GTGCAATGAT AGTCACTTCT AGTTTGCCAT TTCACAAACT 1020
 CTGNACAGCA AGGTATTGGT AGGTACTTCA ATTTCAAAAG GGGCCCATGG CCAAATATGT 1080
 TTAGGAACCG CTGTTTGNAT TTCTTTTTTT GGAGACGCAT TGTATATAAT ATATGTCAAA 1140
 GGCTTTCCGA ATTCTTCAG GAAAGAAATC AGCTTTGTTA AATCCNAAAA AAAAAAAAAA 1200
 AAAAAAATAG ACTCG 1216

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 478 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ATTTCTTATG ACATGGGGGT TGAATTGGT TGGCAAATCT TTAATTTTAA TATCCATAAT 60
 CAGTGAGGTG CTGCTGGCTG TAATCAITAA TTGTGAAATC TAAGGAGCTT AGTTCATGGC 120
 TCTAGAATTT CACAGAAAAR TGYMTATGA TACGAGCATT AAGTTTATTT CTTCTGATCT 180
 TTSATGCAGC TTTGTTGAGT TTATCTGTTT TTGTATTTAT TGGTCACTA CTCCCATGC 240
 CAAAAGGGAC TGGTCTACAT ACCTGCGCTA AACACCTGAT CAAATCACTA AAAGAAAATG 300
 TGTACCTCT AATGAATTAT CCGATTCTA AGTTAAAAAT CAATATTTCC CCGTAGTGAG 360
 GTTTGCTTTT TAAAAAGAAK KCTTAAAAAA AAAAAAAAAA AAACGAGTTN AAGAAAAGGA 420
 AGCAAGCTCA GGTAAGGTGC ATACATGGG CTAAGGAAGC TAGAGCTGT GAGAGACC 478

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 518 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double

TTTTGGGT TGAATTTGA TACTAGTGA TGGCAAGAA TTCAAAAGG ATATATATGA 120

5 GCTACTAGGT AAGCCTTCTG GGACTTTTAC ATATTTTGGG GAAGATTGAT TTTTCTTCTT 180
 ACATCCTGTG GACCTTTGGC CATCAAAATGG TATGGGGAAG CTCATCCGTC TGTCTGTGAT 240
 GGTCAATGCA GTCAGGCGTC TTTTATGAT TACTTGGGTC CTCAGTACTG TCCAGATGC 300
 TTTGGGAGC CGTAGTGTG TGGAGGAGGA GTGCTCAGA GGAATCTGCT GTGTCCAGG 360
 10 CCAGCATAAA CAAGCCAAGG GGAAAAGGCA GGCATGGAAT AAAGGGGAG AATACAGTG 420
 TGTGACTTAC TGTGACTGT GTGGATTAGC CTATCAGCAG TAATCAAGCA GGGGGAGCG 480
 15 CATTATTTT GAGTCAGAAG AGTGAGCACT GGCCGAGGG TCGAGCATCA AGAGGCGGTG 540
 TAGGACNCA AGCCTTCTTN CNGGGAGAC AAGCTCAATA ACCNGTCAGT AGTCACCGAC 600
 AATTTTGGGA AGCAAGGG 618

(2) INFORMATION FOR SEQ ID NO: 62:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 751 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TCGACCCAGG CTTCCGAGGA GCTGCACTTC TGAGACAGCC ATTCTCTTTC CATAGCACTG 60
 35 TCTGTCTGTA CAGCTCATAG AATCAACAA TTTCTTCAA CACTGCTAGG CAGCCTCTAA 120
 ATGGCCCTGA TCACTCTAC CTCTGCTAT TCACACCNNT GTAAAATTCC ACCCCTGGAC 180
 40 CTAGTGACTC AATTCTAACA ANGAGAATAC AGCAAAAGTA ACATCGCTTC TGAGGTGAGG 240
 CTACAAGGAG ACTAGGATGC CTGCTTTGCT CACTCTTCTC CTGCTCTTTC CATTGCTCCC 300
 TCTGATGGAA GGCAGTTGCC ATGTGATGAG GTGCCCTATG GAGAGGCCCA CGTGACAAGG 360
 45 TATTGTAAAA AGCCTCTGAC CAATAGCCAT CTAGAAAGGG AGGCCAGTC CAGCAGCCTC 420
 TGAGATGAAT CTTCCCAACC TGAGTTTGA GACAGATTCT CTCCCTATCC TGCCTTGGGA 480
 50 TGATCAGAGC CACCAACAAAC ACCTTCAGTG CCTGCTGAGA GGCAAGCCA GTGAACCCAA 540
 GGTAAACTGG ACAGAATCTT GACCCACAGA AACTGAGATA ATGTTTGTTA TTTTAACTG 600
 CTCAGTTTCT TACAGAGCAA TAGATAACTA ACTCAAACAC CATAAAATTC TAATATTTTA 660
 55 TTCTATCACA CAAACAGGT AATACCAAGT AATGCCATT ACTATACACA TATTTTGTGA 720
 ACACAATTAC ATGTGATTTT TTAAGAAAGC T 751

60

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 780 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CNENKAMTCA CTGTCCCCGA TCCCCGGCTC GACCCACGAG TCCGGGTTTG CAACTCCTGA 60
 GGGCTGCATG CTGACTTCA CATTTCCTA CCTTCCTTC TAATCTCTTC TAGAGCACCT 120
 15 GGTATCCCA ACTCTAGAC CTGCTCCAAA CTATGACTA GATAGAATT TGATCCCTA 180
 ACTCACTGTC TGGGTCTTC ATTGCTCTA ACACATTTC CTGTGCTTC CTCTAGGG 240
 20 CAGCATCTA AGGAGGGAG GTCTAATTC AACTGGAGA AGCCTCAGTG GTAGAATTTC 300
 AGGCACTTC ACTTCAAGC TGGCAAGGC CAGGATTCG CGAATGAGC TGGGGCTTAG 360
 CTGGAGGTC GTCTAGCA GACAGGGAAT GGGAGAGGAG GATCGGAACT AGACAGTGGC 420
 25 TGGTATGCT CTGAGCTTC CTGGGGCTTC CTCAGCTTC TCTGCTCTT TGTGTTTTT 480
 TGTGATTTG GGGCTTGGG ATCCCTTTG TCTCATCTG AGACTGAAAT GTGGGATCC 540
 30 AGGATGCTT TCTTCTCTT TACCTTCTT CCTCAGCTT GCACCTCTA TCTGGAACC 600
 TGTCTCTCT TCTTCCCAA CTATGCATCT GTTGTCTGT CCTCTGCAA GGCCAGCCAG 660
 CTGGGAGTA GAGGAAAT AAACAGATT TCTATGCCA AAAAAAAAAA AAAAAAACC 720
 35 GCGGCGAAA GTTATTTTC CTTAAGTAA GGGTTAATT TTTAGCTTG GCACTNGGCC 780

40

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TTCCGAATTA ATGACTTAC TATAGGAAT GGGCTCGCCA TGACCCCGCG TAACCAAGGT 60
 GAGCTCGCTT CTCTAGAGAA TATGAAAAAG CAGAGCGACT CGGTAAATG AAAGCTGCGA 120

60

322

AGTCTCACA GGTCCGAGCA CCGATGGCAT TCCTTTGCC CTGAGTCTGC AGCGGGTCCC 360
 TTTTGTGCTT CCTTCCCTC AGTASCCTC TCTCCCTG GGGCACTCCC GGGGGTGAGG 420
 5 GGGTTACCCC TTCCAGTGT TTTTATTCC TGTGGGGCTC ACCCCAAAGT ATTAAAAGTA 480
 GCTTTGTAAT TCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 540
 10 AAAAAAAAAA AAAAAAAAAA AAAANNCGG GGGGGGCCCC CCCCCCCC 588

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 774 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTTAAAGATG AACAAATGAC AAGGGAGGGA GATGAGATGG AAAGGTGTTT GGAAGAGATA 60
 25 AGGGGTCTRA GAAAGAAATT TAGGGCTCTG CATTCTAACC ATAAGCATTC TCGGGACCGT 120
 CCTTATCCCA TTAAATTAAT TTCTCTGACA ATTCAATTAT TTTCTGTTAT TAATGTTGCC 180
 30 ACTGCTTTCT GTTGTCTGTC ACTTCTTGA TAAATATTTG CTATCGTTTT ACTCCAGTCA 240
 TTCGATGTTG CTGAGATTTA CATATGATTC TTGTCAACAT CTGATTTTTT GACCCAATCT 300
 TATTCATTTA ATAAGAGGTC TCATTEATTT GCATGGAAAA ATGCTCATTG TATATTGCAA 360
 35 AGTGAAAATA ACGAGTTGCA AAACATGTA TACATATATG TGTGTATATA TGTACATTTT 420
 ATTTGTACAT TTCTATGTA CATAATGCAA AGGAAAGTGT CTGATTTTAT TATACACCAA 480
 40 AGGTTAACAG TGAATCTCTG TGTGATCTCT TTTTTTTTCT TTTTGCCTAT CTGCATCTTC 540
 TCACTTGCCA AAAAATGAAT ATATGTTTAT GTGTGTATAT TACTTGTGTC ACAAAAAACC 600
 CTAAAGTAGA CAGTAAAAA ACTTGTCAAT CGCCTTTGGA AGGCAATGAA ACACTTAATA 660
 45 AACTCTCAAT AACAGAAGCG TAAAAATGAA ATGTAAACCT CCAATTACCT CTGGATCTCT 720
 TAGCCAGAGT AATAAACTGG TAATTATTAC AGATAAAAAA AAAAAAAAAA AANA 774

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1866 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	ACCCACGGCT CCGTCTCTCT TCTTCAGCAC ATGCCAAAGT TGTTCCTCAC GGCTCTGAG	60
5	ACAAGAGCAT CTTCGATGTA GGACAATGGA AGAGTTAGAT GCGTTATTGG AGGAATTGGA	120
	ACGCTCCACC CTTCAGGACA GTGATGAATA TTCCAAACCA GCTCTCTCTT CTCTGATCA	180
10	GCATCCAGA AAGGAGATA ACCTGATGA GACTTGGAG ATGCTTTCTA TTCAGGATAA	240
	CACAAGTCCC TTGCGCGGCG ANTCGTGTAT ACTACCATA TCCAGGAGCT CAATATCTAC	300
	ASTGAAGCCC AAGAGCCAAA GGAATCAGCA CTACCTTTTA AAAGTCAGC AGCTCTGAG	360
15	TTGGATGAGC TCATGCTCA CCGACTGAG ATGCAGGCA AGTTTGCAGT GAGATCAGAT	420
	GCTGCCAAGA AGCACTTACC ATACAAAGAG GATCACAAGG CTTCCCTGGA CTCATGCTT	480
20	GGGGTCTG AGCAGCAATT GAGGAGCTT GGCATTGCA CATTGCCCAA GGGCATTTGT	540
	GCATCTGCG AGAAACCAT TCGTGGGAAG CTGATCCATG CTCTAGGCA ATCATGGAT	600
	CTGAGCATT TTCTGTGTAC TCATTCGAAA GAAGAGATT GCTCCACTCC CTCTTTTAG	660
25	CGGAGTGGCT TGGCTACTG CCCCACGAC TACCACCAAG TTTTCTCTCC AGGCTGTCT	720
	TACTGGCTG CTCCTATCT GATAAAGTG CTGACAGCA TGAACGAGC CTGGACCCA	780
30	GAGCACTTCT TCTGTCTCA CTGGGAGAG GTGTTTGGTG CAGAAGGCTT TCATGAGAAG	840
	GACAAGAAGC CATATGCGG AAAGGATTG TTAGCCATGT TCTACCCAA GTGTGGTGG	900
	TGCAATCGGC CAGTGTGGA AAAGTACCTT TCAGCCATG AAGTGTGTG GACGAGAG	960
35	TCTTTGTTT GTGGGACTG CTCACAGAT TTCTCTAGTG GCTCTTCTT TGAATCGAT	1020
	GGACGTCCAT TCTGTGAGCT CCATTACCAT CACCGCGGG GAAGGCTGTG GATGGGTGT	1080
40	GGCAGCCCA TCACTGGCG TGTATCAGT GCGATGGGT ACAATTTCA TCTGAGGAC	1140
	TTTGTTGTG CTTCTGCTT GACACAGTT TCGAAGGCA TTTTCAGGA GCAGAATGAC	1200
	AAGACTATT GTCAAGCTG CTCGAATAAG CTCTTCCAC TGTAAAGCA ACTGATCAT	1260
45	AGGCTCTCA GATCTCTAT AAAATTTAAA CCAAGAGAG AGAGGAAAG GTAAATTTT	1320
	TGTACTGAC GTCTGTGTA ATACTCTAT AGAAAAAGG AAGGTGATG ACAAATAAG	1380
50	GAATTTCTAG AATTACATG ACTAGCTGA TAATCTATT TTTTAGGCTT CTATACAGT	1440
	AATTCTATAA ATTCTTTTC TCCCTCTTT CTCGAATCA GCACTTGGAG TTATATCTAG	1500
	GTCTTCTAT CTCTCTCTC TACAGATGA TTTTCACTT GCATAATCA TGGCAACAT	1560
60	TATCACTG TCAATTA TATACATG TTATCTT GATCTT TCAATTA	1620

TTTTGTTTTT CAAGAGGAAG TAGATTTTAA CTGGACAACT TTGACTACTG ACATCATTTGA 1800
 TAAATAAACT GGCTTGTGGT TTCAATAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860
 5 AAAAAA 1866

10 (2) INFORMATION FOR SEQ ID NO: 67:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1152 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

20 CTCAAGSATS TAAAGGCTCT GCAGATTTTC GGAGGCCTGT CTCCCAGCAC CTGATGGGAC 50
 ACTTTTGGCC CCACTGTAAA TTCTGGGTGT ATCTCCACT GTATGCTGTC ACCCCAAGG 120
 CAAGCACTGC ATCTGCTTAG TGAAGGATTT ATTGTTGGGA AGATACATTT TCCCTTKAG 130
 25 CAGAGAGTGG CGTATCCTGG CAGTCTTCGG TGAGCCAGTT GTACCAGGAT TATGAAATGC 240
 AGATGTTTAC TGTGTGATTG TTGCTGTCAI TGCTACTCAG GAGTACTGAC CAGAATCATC 300
 30 TSCAACTTTT AGTTGGGAGA GAGGACCACT ATGGCGGGTA GCTCTTTTCT TTCTTGCCAT 360
 TGTGGGGATG ATTCCAGGCC AAAGATGATG GAAAGSTATG GAAATCATCT GAAAGGTTGA 420
 AGCTTGGCAC CTGAAGGCAT TCATGACTTT STAAGSCAGT TTTCTGAAG GCCAGTTCTG 480
 35 CCCTGGGAGG GACGGAGGTG AATCTCCTG AGTACCTGTG GTTTTCTTAC TTCTGCTGA 540
 ATTTACCTAA GTGCCTGTTG TTTCTTGCT GTGGAGGCTT TCTGGTATTT CATTTGAGGT 600
 40 GCAGATGCCT TCACTTTCCG ACCRAAAAAA CCCCMAACCA ACCTAAGACC TTAATGCAAC 660
 TAAGTYTNCC AAGTACTTTT TAACCCAATG GGATGAACAG CCTGTGGTCT GCTCAGATCA 720
 CCCTGACTGC GTGTGAGAA GEMTNGGCTT TGCCAGGAAA TCCAGGAAG CAGGGCCGGG 780
 45 CTGTGTTGGA AGTGGGTTA GCTGGTGGGG CAGCCTTATT TCAATTAAAA GGGCATGAC 840
 TGGGAGCAGC ACTCCTGAG TTTGTTGCAT TTCTATTGC CCTCAAAATG AGAAACCAGG 900
 50 AAAATAGCAG ATTGGAGCCT TCGAGAAGGC ACTAAATGGC TGTMTTATT GACAAAAGGA 960
 AAACATTTTA CTCCATCTC ACTGATGGCA TCTCACTGAC TTAAATGAA GGCANGTTGT 1020
 AGTAAAAAAA AAGTCTACA TTTTCCACC GGCACGTTCT TATATCTGT TTGTCAGTCA 1080
 55 CTGCTCANAA GGGCATGTTG TTTGCGGAN TANAGGCGCT CTCCTCCCT CGTTTCCCT 1140
 ATAGGTTGGG TG 1152

(2) INFORMATION FOR SEQ ID NO: 68:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2433 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

AGCAGGCGGT GCGTGGGGG CGGAGGAGG GCGKAGCCCG GCTGGGCCAC ACCGATCGGC 60
 15 CCGCCGCATG GCTCCCTGCG AAAGCGTCGA GATCCCGGCC GGGGGCAGG AGGCGTACCA 120
 CTTCTGCGG GTACAAGAAA ATTCCGACG ACACAGAGCT GGTTCGGAGC CTTCTTTGA 180
 20 TTTATTGTT TCTATTAAATG GTTCAAGATT AAATAAAGAC AATGACACTT TTAAGGATCT 240
 GGTGAAASCA AACGTTGAAA AGCTGTGAAA GATGCTTATC TATAGCAGCA AAACATTGGA 300
 AATGCGAGAG ACCTCACTCA CACCAAGTAA CCTGTGGGGT GGGCAGGCT TATTGGGAGT 360
 25 GAGCATTCGT TTCGCGAGCT TTGATGGGGT AAATGAAAAT GTTGGCATG TGCTGGAGGT 420
 GGAATCAAAT TCTCCTGCG CACTGGCAGG TCTTAGACCA CACAGTGATT ATATAATTGG 480
 30 AGCAGATACA GTCATGAATG AGTCTGAAGA TCTATTCAGC CTTATCGAAA CACATGAAGC 540
 AAAACCATTC AAAGTGTATG TGTACAACAC AGACACTGAT AACTGTGAG AAGTGATTAT 600
 TACACCAAAT TCTGCATGGG GTCGAGAGG CAGCCTAGGA TGTGGCATTG GATATGCTTA 660
 35 TTTCGATCGA ATACCTACAC GCGCATTTGA GGAAGGAAAG AAAATTCTCT TTCCAGGACA 720
 AATGCGTGGT ACACCTATTA CAGGCTTTAA AGATGGGTTT ACAGAGGTCT AGGTGTCTCT 780
 40 AGTTAATGCC CGGTCTTTGT CAGCACCAGG AACTACAGGA ATTGAACAGA GTCTGACTGG 840
 ACTTTCTATT AGCTCAACTC CACCAGCTGT CAGTAGTGTT CTCAGTACAG GTGTACCAAC 900
 AGTACGCTTA TTCCGAGCAC AASTAAACCA GTCCCTCACT TCTGTGCCAC CAATGAATCC 960
 45 AGCTAGTACA TTACGAGTC TGAATGCTTT ACCAGCAGGA CTGCCCAAGC TCCCAACCT 1020
 CAAGCTCAAC CTCCAGGAC CAGACATCAT GCGAGGCTT GGTTCAGAG AACTGTGAAA 1080
 CCGAGTCTG CAGCTCTTC CTTCATGCC TCCCGGAAAC TTACCTGGA TTGCACTCT 1140
 50 CCGCTGCGA TCCGATTC TCCGTCATT CCGCTGCTT CCAGAGAGCT CTTCTGAGC 1200
 AAGTCAGGA GAGTCTCTT CTTCCTGCG GCGCAGGAC AACGAGCTT CTGACCTGCG 1260
 60 GCTTTCTCT AGGATCTG ATGAAATCT TCTAGTCA CTTAACTTT TAACTATCT 1320

TTGGAATTGG CTTGGTATAT TTAACCACGG GAGGCTGTCT CGAAACGCAA ACTATCATT 1500
 ATTTCTACT ACTTGTACG GTATCTGTAG GCATCTGTGA AATAATTCCA AGGGGAAAAC 1560
 5 TAAACGAGGA CTTGGTTGT ATCCTGCCAG GTTGAAGTGG GCTCACACGC TAGGGTGAGA 1620
 TGTCAGAAAG CTTTGTATT TTAACAACG AAAAAGAAAT GTAAGGTTGG CTTCTTGCCA 1680
 GGCTTGCACT GCTTTTCTG GGGGTGTGCA TCTTGGGAA AGGTGGTGGC GGGGCTCCA 1740
 10 CTAGTTTCC TGTCCCTGC TGCTCTTCC GTAAGAAAT GAAATATTCT ATGCTTAATA 1800
 CTCACACGCA ACATTTCTT TACTTTGTAA GTCTTTTGG AGAATGAGA CCACCTCACT 1860
 15 AAACGTAAA CCGTAAAGAG ATTTTACTT TTGCTCTCG TGAGTGGCAT CTCTACTAAG 1920
 GTTTACACAG GAATTCACG TGAAGACTG TGTAAAGTT CTACAGGGG CATTGTTAAC 1980
 TGAACGTCTT TTTTTCAGC CTATACGGG ATCTTGTCT TGAGCTCTCA GAATCACTCA 2040
 20 GACAACATTT TTAAGTGGT GTGTGTGCT TCTACATACA CTTATAAAG TGACATTTCA 2100
 AAAGAAATAA GTTCCACAG TTTTAAACCA GAAGGTGCA CTCGTGGGT CCTGTAGTA 2160
 25 TTATAGCTAT ACTGGGAAA CATAGATACA GCAATAAAGT ACAGTAATTT TACTTTTTTT 2220
 CTTGTGTAC ATCTAAATTA CAACCTTAA TTGCCACGT TGCCTTACT ACTCTCCAGT 2280
 ATGTCTTAT ACTCTCCAGT ATGTCACGCA TCTTTAAGTT TGCAGTCT ATGTTTGCTT 2340
 30 TCTCCCATTT TTAAGAGATG GTAAGTTAAC TGGAAATGAT TACTGAATG AAATTAATG 2400
 CAGATATCCC TGTTTTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2460
 35 AAAAAAAAAA AAAAAAAAAA AAA 2483

40 (2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

50 GAGAAATGGA GCTTGTAG AAAAAATTT TTTCAACGCA AACAGTCATT TTCCAGTGAA 60
 AGGAGAGCGT ATCCGCCGTA GGATGGACTT AGATCGTGA AAAGCTGAGG CCACCGAGGA 120
 TATAACCTCC GGGTCTTTT GCCTCTTTT CCTTAGACTC CCTCCAACT CGTGTATCTT 180
 55 TCCTTCAGCA GTACTGGGCT CCACGCGAAC CTAGTCTTT GTCTTTACCC TATTACCTTT 240
 CATAACATCC TAGTTGAAAA GTATTATTC AACCGCTTT GAAAATGAGA ACAGTTTAC 300
 60 AGARGCTAGG TACTTGCGA AGGTCGTTCA ATTAGTAACC AGTAACGCCA GGACTGCCAG 360

TTTCTTGCTT CCGAATTCTC ATGGTAGCTT TCACCARGCT CCCCCTCMAA TGCTAACGTC 420
 AACTACTGAA CTAGATTAGC AAAAAGGTCT TTTAACAGAA TTCTTGTTT TCACAGAGAG 480
 5 TTTCTTTCAT GAAGCGCCCC ATTTCTACAG AGGAAAATAA ACTCCAAGCA CCCAGT 536

10

(2) INFORMATION FOR SEQ ID NO: 70:

(1) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCACGCGTCC GGCCTTTCTT TCCAGAGGC GCCGGTTGGA CTCACGGGCG GAGCATGATG 60
 GSTAACAGGA CCGGTGGGT CCCCAGGAAG TCTAGAGGG GGTGGGGTT TGGTGGACA 120
 25 AGCTTTCCTC GTCTCTCCC GACAGAGCTG AGTGTCTCT GGTTCACCG GAGGGGGCA 180
 TTCCACCGG ACBSAGGCT TGGGGTGT CCGGGCTGGG GAATAGTAG GGGTTGCCG 240
 GCGGTGTGG GATTGGGGC GTGTGGCTG AGTCCCGGA GTTCTTGGAG GAGGTGGCC 300
 30 CACCGAGCTT CCGGACCGC TGATCTGCC GTAGCTTGC GGANGGARG CTGAGCTGAC 360
 TCTCCGTCCC TTCTCCATC CCTTCAGTG GTGGGTAGG GCACCTCGCT GJCCTCTC 420
 35 TCTCTCTGT CCTGTGTGT CTTCCTGGG ATGCAGATGT ACAGCCGTCA GTTCCCTCT 480
 ACCGAGTGGC TCACCATCA GGGCGGCTG CTTGGTTGG GTCTCTTCT GTTCTCGCT 540
 ACTGCCCTCA ATAATCTGA GAATCTTGT TTTGGCAAAG GATTCCAAGC AAAGATCTT 600
 40 CCTGAGATTC TCCTGTGCT CCTGTGGCT CTCTTTCAT CTGGCTCAT CCACCGAGTC 660
 TGTGTACCA CTTGTTCAT CTCTCCATG GTTGGTCTGT ACTACATCA CAAGATCTC 720
 45 TCCACCTGT ACCAGGAGC AGCTCCAGT CTCACACAG CCAAGTCAC AGGCAAGAG 780
 AAGAACAGAA ACTGACCTG AATGTCAT AAAGTGAAT CTCTGTAAAA AAAAAAAAAA 840
 50 AAAAAAAAAA AAAAAAAAAA AAAAA 865

60

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 865
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

5 TCATCATATA CAAAGTTTTT CGTCACACTG CAGGGTTGAA ACCAGAAGTT AGTTGCTTTG 60
 AGAACATAAG GTCTTGTGCA AGAGGAGCCC TCGTCTTCTC GTTCCTTCTC GGCACCACCT 120
 GSATCTTTGG GGTCTCTCAT GTTGTGCACG CACAGTGGT TACAGCTTAC CTCTTCACAG 180
 10 TCAGCAATGC TTCCAGGGG ATGTTCAATT TTTTATTCCT GTGTGTTTTA TCTAGAAAGA 240
 TTCAAGAAGA ATATTACAGA TTGTTCAAAA ATGTCCCCTG TTGTTTGGG TGTTTAAGGT 300
 AAACATAGAG AATGGTGGAT AATTACAAAT GCACAAAAAT AAAAATTCCA AGCTGTGGAT 360
 15 GACCAATGTA TAAAAATGAC TCATCAAATT ATCCAATTAT TAACTACTAG ACAAAAAGTA 420
 TTTTAAATCA GTTTTTCTGT TTATGCTATA GGAAGTGTAG ATAATAAGGT AAAATTATGT 480
 20 ATCATATAGA TAACTATGT TTTCTATGT GAAATAGTTC TGTCAAAAAT AGTATTGCAG 540
 ATATTTGGAA AGTAATTGGT TTCTCAGGAG TGATATCACT GCACCCAAAG AAAGATTTTC 600
 25 TTTCTAACAC GAGAAGTATA TGAATGTCT GAAGGAAACC ACTGGCTTGA TATTTCTGTG 660
 ACTCGTGTG CCTTTGAAAC TAGTCCCCTA CCACCTCGGT AATGAGCTCC ATTACAGAAA 720
 GTGGAACATA AGAGAATGAA GGGGCAGAAT ATCAAACAGT GAAAAGGGAA TGATAAGATG 780
 30 TATTTTGAAT GAAGTGTTTT TTCTGTAGAC TAGCTGAGAA ATTGTGACA TAAAATAAAG 840
 AATTGAAGAA ACACATTTTA CCATTTAAAA AAAAAAAAAA ACTNGAGGGG GGCCCGGTAC 900
 35 CCAAAATCGCC GCATAGTGAT CGTAAACAAT CT 932

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 996 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

50 CGCCTGGCAC CATGAGGACG CCTGGGCGTC TGCCTGTGCT GCTGCTGCTC CTGGCGGGAG 60
 CCCCCCGCGC GCGGCCCACT CCCCCGACCT GCTACTCCCG CATGCGGGCC CTGAGCCAGG 120
 AGATCACCCG CGACTTCAAC CTCTGCAGG TCTCGGAGCC CTCGGAGCCA TGTGTGAGAT 180
 55 ACCTGCCCCAG GCTGTACCTG GACATACACA ATTACTGTGT GCTGGACAAG CTGCGGGACT 240
 TTGTGGCCTC GCCCCCGTGT TGGAAAGTGG CCCAGGTAGA TTCCTTGAAG GACAAAGCAC 300
 60 GGAAGCTGTA CACCATCATG AACTCGTTCT GCAGGAGAGA TTTGGTATTC CTGTTGGATG 360

ACTGCAATGC CTTGSAATAC CCAATCCCAG TGACTACGGT CCTGCCAGAT CGTCAGCGCT 420
 AAGGGAACTG AGACCAGAGA AAGAACCCAA GAGAACTAAA GTTATGTAG CTACCCAGAC 480
 5 TTAATGGGCT AGAGCCATGA CCTCAGAGG TCTGTGTTA GTGTATCTG AACTGTTAT 540
 GTATCTCTCT ACCTTCTGGA AAACAGGCGT GGTATTCCTA CCGNGGAACC TCGTTGAGC 600
 ATAGAGTTAG CAACCATGCT TCTATTCCC TTGACTCATG TCTGCCAGG ATGTTAGAT 660
 10 ACACAGCATG TTGATTTGCT CACCTAAAAA GAAGAAAAGG ACTAACAGC TTEACTTTTA 720
 TGAACAACCTA TTTGAGAAC ATGCACAATA GATCTTTTTT ATTACTCGTT TAAATGGAGTA 780
 15 ATGGTACTTT TATTCTTTCT TGATAGAAAC CTGCTTACAT TTAACCAAGT TTCTATTATG 840
 CTTTTTCTA ACACAGACTT TCTTCACTGT CTTTCATTTA AAAAGAAATT AATGCTCTTA 900
 AGATATATAT TTATAGTAGT GGTGACAGGA CCCACTCTTT CATGAAAGG TGATGAAAT 960
 20 CAAATAAAGA ATCTCTTCAC ATGARAAAAA AAAAAA 996

25

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 785 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

35

GGCACGAGCG CCTTTGCGTA CACAATAGCT GCTAGGAGTA CCAAAACCT GATACARCC 60
 TGCTGGTCTC ATGCCACGT GTGAGCAGGC CAGCGTCAMA CGCTCGCTG TGACCCGTCC 120
 40 CGRAGACTGA AATGGGCTG GGTCTTCTCC TKGTCCTGTG ATWAAATCC TCTCTGAAA 180
 GTGGAGAGCA AAGACAGACA GAGGTGCGG CTCACAAGAA TTCTCTCTG TACTGGTA 240
 ATCAATGTTA CTGTGTTTC CTTTCAGGA AAGACCACAG CAGATTTCT TCAATCGTCT 300
 45 CCTCCTAGCC TGAGGAGCA GCTTGAAGT GACCTGAC ATCAAAGAG GATATATGTG 360
 GCTGCTAAAG CCAAGGCTC ACAGCCCTGT TCACRTCTG GTCTTCTCT TTCCAGAGG 420
 50 CTGGTCCCAG CCAGGCACAC ACAAAGGCA GATTCTGTA AACSCAGCT CCCTCCCTGG 480
 AGGCTGCCTC CTGCCCCGA TCTGGAGTGG AGCTGCTCTG AGATTTGAG TTCTCTGCA 540

60

TAAATGTAT GTATTTTTT TTTGAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 780

AAAAA

785

5

(2) INFORMATION FOR SEQ ID NO: 74:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1069 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TCCTCACCAT TCCCTAGG CAGGTCCCTG CAGGTCCCAC ACTTCTCCCA GGTCCCTAAA 60
CTTGGGTGGG TCTTTCCCT GGAGTAGCTG GNTCCTCCAG TCGAGGTCC TGTTCAGTCG 120
GTTCTTAGGC TCTTGCACAT GAAGGTGTGT GCCTGTGGTG TGTGGGCTGC TCTAGGAGCA 180
GATACAGGCT GGTATAGAG; ATGCAGAAAG GTAGGGCAGT ATGTTTAAAT CCAGACTTGG 240
CACATGGCTA GGGATACTGC TCACTAGCTG TGGAGGTCCT CAGGAGTGA GAGAATGAGT 300
AGGAGGGCAG AAGCTTCCAT TTTTGTCCCT CTAAGACCC TGTATTCTT GTTATTTCCT 360
GCTTTTCCGA GTCTGCACT GGGCTGCCCT GTACCCTGAA CCTCATGAGC CTCTAAGGGA 420
AAGGAGGAAC AATTAGGAGG TGGCAATGAG ACCTGGCAGG CCAGARTACA AGCCCAGCAC 480
CAGTGTCCCA GCTTACTGG GTCCCTACCC TGGGCCAAAC AGGGAGGGCT GATACCTCCT 540
TGCTCTTCCT AGATGCCCACT CTCCTACAAT CTCAGCCAC AAGTCCTCTC CACCCTAGGG 600
GGCTTGCTGC ATGGCAATAA CTCATAATCT GATTGGAGG TTTGCCCTTT ACAGGGGCAG 660
ATTTTCTGCT CATTCAACA ATGAAATGAA GAGGAACTCC CTCTTTCTAC AGCTCACTTC 720
TATCAGAGGC CCAGGTGCCT CAGAGCCACA TTGAGTTGCT TTTTCTGGGA TGAGGAAGTA 780
GGGTAAACT CCCAGTTTC CTGAGGGAGG CTCCTGACAG GTGCCCTTTG TCAGACCCTA 840
CCACAGCCTG GATAGGCAGC CACATTGGTC CTCGCCCTTG CTCGGNACTC CGTGGTGGTC 900
CTGCCCTTCT CCTGCATGC CTGTGGGTCT GCTCTGGTGT GTGAAGGTGC GTGGGTAAAC 960
TGTGTGCCTA CTGAACCTCG CAAATAAACA TCACCCTGCA AAGCCAAAAA AAAAAAAAAA 1020
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1069

55

(2) INFORMATION FOR SEQ ID NO: 75:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 831 base pairs

(B) TYPE: nucleic acid

	CTTTTGTGTC TGTAAAGATAT ATGCAGCCTC ACAGAAGCAG CTCTGCTGCTC CACTTTACCA	480
	GCTACGTTTTT TATCTTAAGC ACATGCGGCT CCGTTAGAAG TTAAGTCACT GATTTAAAAA	540
5	AAAAAAAAA AAACCTGAGG GGGGGCCCCG TAACCATTCG CCGTAAAAAT	540
10	(2) INFORMATION FOR SEQ ID NO: 77:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1274 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:	
20	GAGGCAACAC ACCTGGGCTG GAAGGAACCT CTTAAATCA GTTTAGCTCT TGTATTTTGT	60
	TCTGTGATGG AGGACACTGG AGAGAGTTGC TATTCAGTGG AATGATGTCG AGTCACTGGA	120
	CTCTGAAAAAT CTTATGCTT CTTTATTTT ATTTAGCTTT AGAGTTGCTT TCTGGGTTTG	180
25	TATTATCTCT GGCAAAATGAC CTGGGTTATC ACTTTTCTTC CAGGGTTAGA TCTAGATCT	240
	TGGAAATGCT TTAGAGAGCA TTTTGTCTCT ACCAAGGATC AGATACTGGA GCGCCACATA	300
30	ATAGATTTCA TTCACTCTTA GCTACATAG AGCTTTCTGT TCTGTCTCT TGCATGAC	360
	TTGTGCGGTC ATTACACACT TACAGTACC AGGAGACAAA TCACTTACAG ATCCCCGAC	420
	ATGCTCTCTC CCGTTGGCAA GCTCACTTGC CTTGATAGTA GCACTTTTCT GTTTCTGATG	480
35	TACCTTTTTT CTCTCTCTCT TTGATCAGC CAATTGCTAG AATTTGCTCA GGCATTTTGT	540
	AGAGGACCTT TTTGGGCTCC TATATGAGCC ATGCTCTCAA AGCTTTTAAA CTTCTCTGCT	600
40	CTCTACAAAT ATTCACTACA TCACTACTCT CATCTAGAA GCTTCTGAA AAGAGGGGCA	660
	AGAGCCACTC TCGCCACAAA AGGTGGGGT CCACTCTCTC TCGAGGTTG TGAAGTTT	720
	CAAAATGTAC TAATAGGCTG GGGCCTGAC TTGGCTCTGG GTTTGGGAG GGGTAAGCTG	780
45	CTTTCTAGAT CTCTCCACT GAGGCATGGA GGTGTTCTG AATTTTGTCT ACCTCAGG	840
	GATGTTGTA GCTTGAAAA GGTCAAAAA TATGGCCCC TTGAGCTCTT TGTAAAGAAAG	900
50	GTAGATGAAA TATCGGATCT AATCTGAAAA AAAGATAAAA TGTGACTTCC CTTGCTCTGT	960
	GCAGCAGTGG GGTGGATGC TCTGTGGCT TTCTGGGTG CTTATGCCAC CCCACAGCTC	1020
	CCAGGAACCT TGAAGCCAAT CTGGGGACT TTCAGATCTT TACAAAGAG GTACCAGGCA	1080
55	AATCTCTGC TACACATGCC CTGAATGAAT TGCTAAATTT CAAAGGAAAT GGACCTGCT	1140
	TTTAAGGATG TACAAAAGTA TGTCTGCATC GATGTCTGTA CTGTAAATTT CTAATTTATC	1200
60	ACTGTACAAA GAAAACCCCT TGCTATTTAA TTTTGTATTA AAGGAAAATA AAGTTTGT	1250

TGT TAAAAAA AAAA

1274

5

(2) INFORMATION FOR SEQ ID NO: 78:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1133 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AGGATTTTTC CTTCTTCAAC CAAAATCTGA GCATTCTTTC TATGTTGAAA ACACTGAAAA 60
 ACTAATTTWA GTTAATGAAC TAGAAAGAAT ATTGATTTTW AAGAAACAGA AAAATACTAC 120
 TTATTTTCCT TCTAAATAA CGTTTCTTTC AAAAAGTCTCT GGCTGAAGTA TAACATGCTG 180
 GTAGTTAACA TAAATCTTGT CTTCTCTTTC TTCTTTATCT TTCTTTGTTA TTTAGATGCT 240
 TGTATAAATG TCTTTTCTTT TTATTAAGTG CCTAATTGAC AGAGCTTAAT TTGAAGAAGT 300
 GGCCTAATTT ATTACCACT TAAGAATTGC CTTTATTGGG GTATTTTATT TGTTCCTGCG 360
 TCTTTTTCAT GTTCTTCAGT CTATTCATCC CTGTGAGTAT GTGTGGGGGA CAGCTGATAG 420
 AAGCGAGGAG AGTGTGTCTA TGCTCAGGAT TCCCTTTTAC CCACTCAGCC AGAGATCCAC 480
 AAGGAGCAAC AAGGACAATT TCACATGCTT AGACTTTCTT GGAAGAAACA GTGAGGAGGA 540
 GTAACTCTTG AGTAGTCTCA AGCTGCGATG ABAATTGTCC TAAGGAGATT GACCCACCT 600
 TCCAACATCT TTTCACTTTA TTTGCCCTTC CTTACATTTT GGTTAGGTTC CATTTGGATT 660
 TGCAGCAATA ATGACTTTAT TTCTCTCTTG GTCAGGATTT GGCACATAAA ATCCTTTTAT 720
 TATAGAACTA CTTATTTTAC TTACATAGTA ATGTAAGTAA TGGAGAGATT TATAGAGAAT 780
 TTGKTCTTTC CTTCATATA TGTCATTTT GAGACAGAT ATGATAGAAC TAGAAATTAA 840
 GTTGCATTTT TGAATTTTCT ATTTGAATGA ACTTCAAGTA TCTTCTTAAT TATTAAATTT 900
 TCTGATGAAG GCTTCTTAAC AAATATATAG TATTATTAAA TCTAATTAA ATTTCGAAAT 960
 ATTAATAAAT AGGTATTTTA TTACTGTAA AAGTCAAA TCCATTATGT AGATAAATCT 1020
 TATTCTTTTC ATTCTTTTCC CTGTTTACAT CTTTTTACAA AAGCTTAGTC ACCAATTAAA 1080
 CCTTCTCTAT CAAAAA AAAA AAAA ACTCGAGACT AGTCTCTCT CCT 1133

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 661 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

	GAATTCGGCA CGAGGGGAAA AGGATGCTGA ACGAGAGCAG AAAGCTCTTT TCCTTTGCTT	60
10	CACGCTTTTC CAGTCTTTAT TTAAACTCG GGTTCCTTT CTGTGGTGGC AGCAAGCTTT	120
	ACTCCACCTG CACTGCTGCT CCGGGGGGCT CCGCAGGCTT CCTCTGCTT TTCTACCCAG	180
15	TGGCTGAGGG GATGCTGTG TTGCTTGGAC GCAACACTGA TCTCTGTTC CTCACCTTGG	240
	CTTTTCTCTT GGGTCTCTCT GGGGTGGAAG CTGGTCCATG TGTCCCGGAG AGTCATGGCT	300
	GCTCTCTCTG GCAAGGCTCT GTGTGCTGA CTCTTCCAC ACTTGGGGG AGCTGGGAG	360
20	CCCGTCTCTT GTTCTCTCTG GTGTCTTGGT ACAGAGTTCG AGCTGGGAG TCTCCGTGGA	420
	CCGAGACTGG GCAATTTTGGC AGGGGGGCGA TGGGAGGAGC AGTGTCTTGG CCTGGGGGCT	480
25	GTGTCTGAT TTGTGAGGC CCGAGAGCAC AGAATTTGCT GGCACCTTGA GGTCTTCTCT	540
	GGCATGTCTC AGATTACATG AGTACGGCT GGGAAATATG TTTCTTTTTC GTAATGAGG	600
	CGTCTTTTAC ATATAGTAAA GCTCACCAAA AAGTAAAAA AAAAAAAAAA AAAAACTCG	660
30	A	661

35

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1378 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

45	ATTGGGTACC GGGCCCCCCC TCGAAGTTT TTTTMTTTT TTTTAATGAA AGCTCTCAAA	60
	TAAGCGATTT TATTCTATC CATGATTGCA GACATTTACA AAACCATAAC ATCTGAGTTC	120
50	ACCTTAAAAA ATAATTATA TAAAGCAGTG ATATACACAG CACAAAATAG TTCAAGGAGG	180
	GGCAGGAGC AACTTGTAAT AATTAAAATG TAAACGTGAA AAAAAGGATG GAATAAAAGT	240
	CCCTACTTAT TTCTACTTAA GATGTGATGT GATAATATT TACAATGTCT TGTGGGTCAA	300
55	TGTATGTATG TGTATATGTC TGTATAACAT ACACATATAC AGTACATTCT CTTTCCACAC	360
	CATATACATA CACACATAAT TATTTGCAST TCAATTTAGG GCAATTCTAA TATGCCACTC	420
60	CGTACAGTTG TTTGAATCAC ATTTGGACCC GCTTCTTCA CAAAAGAGGG GAGAGAGCAG	480

335

GAAATAAAAA GGTGGGTTTG GTGTGACTGA GATTCCCTTG TTAACTGTA CACTGTGATG 540
 AATAATTTTC TTCGATAGTA GTTGTGTGAA GGGCTGACTC ACTGTGGTTT TCATGAGGAG 600
 5 ACTTGGTAAT GGATCAGACG CTCATTGTCA TGCTAGGGGA GTAACCTCA CTCTGAAAAG 660
 GATTTAAGAA ATTTGCGCCG ATTTCGUCAT CATTCCCTTG AGTGGCCGGT TGATTACTCA 720
 GCTCATATT ATTGGGAGAA TTCTTGGAAA TACTGTCCAT ATCTCCTGAG CCTAAAGAGC 780
 10 CATTGATGTG ATGTGACTCC ATTCTCTCTA ATCCACCCAT GGGACCATCT GACCCAGGRC 840
 CCATTGGAAA ATTAGGCTCG TTAGGTCCAG GAGGTACTGC ATTCAATAAA GTATACATCT 900
 15 TATCACCAGA GTTGGTGGAA TCTGTGTGAG TAGGATGAT GGGTGTTCCT GTTGGCCCTC 960
 CACTCTCTCG AGGACCTACA TAATTCGCCG GAGATGCTGA GGAGTATGCT ATTGAATTGG 1020
 CATTTGTTCG GTTGGGCAAA GGTCTACGAC CACTGGACG CATGTTGATT CCAGGCATTC 1080
 20 CAGGGCCATC TAAAGCAATC AGTGGGGCTC TTATTGCACC TCATAGTTTC TGTGTGCTA 1140
 AGGGCACCAT TCCTTTTGGG GGAGTCATTC TTGCAATTCG CCGACCCATA TTTGGATGTC 1200
 25 CTTGTTGTG AGTGGATCC ATTCCACTGG GGAGTAATGG CTGACTTCTT GGGACACCTC 1260
 CAAGTGCCTG ATTAGGTATC CTCATGGGGG GCCTTGGACC TCAGGGCTAC CGAGGTGACA 1320
 30 TAAAGGGTA ATCATCGAAG GCTTTTCCTT CACTTGAGTG TTCACATGTT TCACGTCT 1378

(2) INFORMATION FOR SEQ ID NO: 81:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

ACTTGTGCA AATGCTGCTG TCAGATGTAG TCAGCTGNAG AAATTTAAAA TGAATTGCA 60
 45 AGTGAAGAGT CTGCGGATTA ATTGGGCTTT AATTAACAGG CTTTATCAAT GTTCTCTCAA 120
 GGGAGAGGTC CAAGCCTAAT TAAGGAGCTA AACTTCTGTA GTGAGGGGCT GTGAGGATGG 180
 50 AGGTGAGGA GGCATCTGGG GGGGCTGCTG GCGGGGCCAG CAGATGGGGC CTGCTCTGGT 240
 GAGCTGCGCG CACCGGCACT TTCTTATTT CCACTCAGTA AGGCAGAGAA GGCAGAGTGA 300
 GAGCTGCGCG CACCGGCACT TTCTTATTT CCACTCAGTA AGGCAGAGAA GGCAGAGTGA 360
 60 CTTTCTCAAT GGTGGGCTG GTTGTGTGAG AATAATTTTA GAAATTTTA AATTAAGAA 340

5 TTTTMMTTA GTTTTACCT TTTCTAATT ACCCTTATTC CGAATGGAGG AACACTTTCT 600
 ACCACTGCTG ACCATTGTAA AATACCCCTGT ATATAAATCC CATTGAAATA ATGCCCTGGA 660
 ATAGAACATC TCAAATGCTG CTTAATTACA GACTACGGTC GATTACTTGT ATTTCATGTA 720
 A'DGTCTCTCC AAGTTAGACA TCTGCTGCAA GACCAACCGG GAGACCATCG AATTGTGAAA 780
 10 AGTACAAACT GACAGTGTGT ATATTTAATT TAAAGATTTA TTFAAAAAT CACAAGCTCT 840
 CACCTAGACT TDCGAGAGCA GTCTGTTTTT TGTAAATGCT GATACTAGAA ACTAATTTGC 900
 TTATTTTAGT TGTATTCAAG ATTGGAAGAT GTATTTTATA GACAAGTTCT GTTTTGAAC 960
 TTTGTGGAAC TGTTCGAATC AATGAATTTT CCAGTTATGA TGAGTATTTA CATTATGAAT 1020
 GTATAACCCA GACATGATTT GTAAATCCGA CATTATGTTT CTATTACATA ACACTTTTIG 1080
 20 ATACAGCGTC TCTTGTCTTC ACTGATCTG GAGTGTCTGT TGTCTGCTG GTGCTTCTGA 1140
 GTTCTGATTT ACAGACACAA TCATCTGTG ATTTTATTTT TAATATGAT ATGCTATGAA 1200
 ACTGTGATAC ACTTATAATT CACTGTCTCT GCATCAGGAG ATGGAGTGG GAAACTGTA 1260
 25 TTTAATACAG TTTGATCTG AATAATCTGT ATGCTTTATA CAGTTTGTCT TGTTCAGAGA 1320
 TGTMTAAAGT TTGATCTTT TTTTCTAAA GATTAAAAAA GCACTTGTCT CACTGTAAAT 1380
 30 ATACAGCATG TAAATTTCT RTAGTATATA AATGGCAGCA AATCACAAAA AAAAAAAAAA 1440

35 (2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1381 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

45 CCGGGGCTGC AGGAATTGK YACGAGGCCA GCAGTTGCTC CCAGTTCAGG AGGTGCTCT 60
 GTACCTTGGC CACAGCCDAA TCCTGCCACT GCTGACATCT GGGGAGACTT TACCAATCT 120
 ACAGGATCAA CTTCCAGDCA GACCCAGDCA GGCACAGGCT GGGTCCAGTT CTGACCTGAG 180
 50 CACGTTTTTT CCTCATGTGA CTTCTGGGAA GCGCTCCCT CATCTGGGC AAAAGGAAGA 240
 GGACGAAGCC CTCTCAGCT GGCCTGTGTT TGGGGCATGA ATCTCTCTCT TCTTCTTGT 300
 55 CTGCTCTCTT TGACAAACCG GGCATGTTTG GCAGTAAAT GGCACCGTGT CACACTGTTT 360
 CCTGGGATTC AAGTATGCAA CCAGAACACA GGAGAAGAAA AGCTCCAGGA TCCCTGTCCC 420
 60 CATCTGTCTT CTTGATGTGA GAGAGACTCT GAGACTTCTT CCATCGCAAT GACCTGTATT 480

337

AAACACAAGC CCCCAAGCA AAAGAGSAGG TTGAGTMTTC TCCAGSATT CAGATCAGCC 540
 CTTCCCAGGS TCTCAGGTG TCACATGATC ACAGTTCAGC GAGAGGCTTT CCGTACCCAC 600
 5 ACTGCTGTA GCAATTCAGT CCACTCTGTC TCAGAGGAG GATTTCTTCC TGATTTTAG 660
 CAGSTTTAGA GCTTCAGCT TCAGTACAA TAGAGGGA AATGGAAGS ATTAGCAGCT 720
 TTTAAAAATG TTTAAATATT TTGCTTTGCT AATGTGTGA TCCCACTAA CTCTCTTTG 780
 10 CAAAAGGAAC TGCTCCCTCG GCTTCCCA GGTGGGCT CTGAAGGAT TCTCACTGT 840
 GGGCAGTGC CTTGACCTC AGGCAGCAAT GTTCATCTCT GTCAGTTGT CTGTTTCCA 900
 15 TGTATTUTAG GCCAGCTAGS CAACACAGAG CCAAGGGGG TTTTGAAGC CAGACGGAAC 960
 AGTGTGGGG CAGGAAGGTG GATGCTGTG TCATGAGCT GTGGAGTTG GCACCTGTG 1020
 TGCTGGTGC CTTCTGGCT CACATSTTCA CAGTGCAGCT CTTGGCAGAC TTGGSTTTTC 1080
 20 TCTTTGCTG TTTTAAAGT GCTTATCTG CAAACAACIT CTTTCTCTT TCAGGAAGT 1140
 TSAATGGCTA CAADAAGGAG CTAGTAAAC TAGAAGTCCA GGTTCCTTG GTTTACTGT 1200
 25 TTATAAGAAA TCTGAAGCA CTTCTACAT TCTTTTATT AACTCACTC TCAGTTGAAA 1260
 GATTTCTTCT TTGAAAGTC AAGACCGTGA ACTGAAAAA GTTTGGGCT TTTGGGGA 1320
 CCAGATTTT AAGATAAAT AAATATTTT ACTTCTGTCA AAAAAAAAA AAAAAATNT 1380
 30 C 1381

35

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1706 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

50

ACTGCACCAC TCCACAGTC TCCCGCTCG ATGAAGAGCT GTCCATGAG GAAGCTGGT 60
 AGCTCAGACT GGAGATAGC TTCAGSAAA AAGACAAGTG CCTAAGGAA ATCAGGCCC 120
 CCAACTATCA TCTGAGGGCT AAAGATGAGA AGTAGATCAC TTAATAAGAC AAAAGCCTGT 180
 AGGGGGAAAA GAAAGGATGT TTAAGGAC AGAATGTTTC CCAAGGTAGA AATGAGCTG 240
 GGGGAGSANT GAAAGATAG AGTGTCTCT TCTATTATG CTATTAAGA CCAAGCTTA 400

60

CCTGGGGAAG CATCTGATTT AGAAATGTGG GTTAGTGTCT AGAGAATGGA AAAATAGACA 540
 AGAGTCAAGG CTGGCAGGAT AACCTGTAAE AACAAAGGGT TTGAAAAATG AGGTTTGGGT 600
 5 TAGGAGAGGG AGAGACAGAT AGGAGAGAAE ACACCACTGA AGAGGAGAGA AAATGAGTAA 660
 AGGGAGAGCT AATTGCTTTT CCACTGGAAA ATGAGTGATA TTCTGGACAT TCTTCAGAG 720
 10 CATCTACAGG AAGTAGAAAT GTACCGCTC CCTAATTTAC TGTACGTCTT CTAGAATCCC 780
 TCAATATTAT CCTTGGCTT CAGGAAATC AAGAAGACCC TGGAGTAGA GTCCACCTTC 840
 TAAGAGAGGA ATGTAAGAGG TGACCCCCAC CCACATGATC TTCTCGCTT TGTCCACTCC 900
 15 ACGCACTGAG ACTTGAACAC CTTAGTGGCC ACCTAGAACG TAGGTCTTA AAATTTAGCC 960
 CCCCACCCC CAACCCATCT CTAGCTCTTC CACTCACTG GTGAGGAAC TTCTGTGTCT 1020
 20 CCACACCTTT CTGCAGGAGT TGGCAACATG GCTCATAGAG CTCCCAGCGA GTCAGTCTAT 1080
 GAGTGTCTTT GGGGAGAAA GGGAAATGTTA TACTGGAAAA GAACAGAGGG AAACAACTTC 1140
 ACAGACACCA GTAAAAACCG GATGGGGAAG AGGAGGAAAG CCACTCACTT GTAGAAGGCA 1200
 25 GAGAGGGCTT TCAGAGTGGT TGGCAGATTA TATAGTCTAT TCTCATCTAG GAAGGACGAT 1260
 TGAGAAAGAA AGAAGATCCA CAATAGCATT TCCCCAGAA CTCATCAGTC CACATCCCCC 1320
 30 GTCTGACAGC CCGTCCGACC CTTGTTTGGG GTGTCCATT GTCCAGCCCC AGCTCCTAGC 1380
 TGTAAAGCTT CTTCAAGCTC CTGCTGGAAR CGGTCACTCA GCAAAATCTAC TAGCTGGCTG 1440
 CCGGCAAAAT CCGCCCGGCT GAAGAAAGTG AATTCGGGAT TACAGAGCAG GTAAGAGCAT 1500
 35 GGGGCTCAGC CTCAGGACC GTGCGCTCTG CATGCTTCAC CACCACCTCC TGAGTTGCT 1560
 GCAGGAACAG CTCAGGTCG TGAGAAGAAA AGGCAGAAGA TGGTGTGCTG TGCGGATGGG 1620
 40 AGGAGGACAC TCTTCTGGCG GGAAGTGGAA CGGGGTAAA AGCATTAAAC TTCAAGGATA 1680
 AGATGCCTAA RAAAAAAAAA AAAAAA 1706

45

(2) INFORMATION FOR SEQ ID NO: 84:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

60

GAATTCGGCA CGAGCTTGGT AGCCTTAGAA CTGCATGAGC TGCTTTACCA CTGGGAAACA 60
 CGAGCACAGC CTAGCTTGAT TTTGTATGTG GTATCAGATC TAAGGTGGAT GGAATTCAGG 120

339

ACTTCCTGTC TACTCTTTGA TTTTGTMTTA TTTTACAAA TSTTTTATTT TSTTTTATTC 180
 ATTTATTCAT CTPLAGAGAC ATGGTCTGGC TCTGTTGCCC AGGATGGAGT GCATGGTGTG 240
 5 ATCATAGGCC ACTGCAGTGT TGAGCTCCCG GGCTCAGGCG ATCCTCCTGC CTCAGCTYCC 300
 TTAGTAGCTG GGAETATAGG CACATGCCCT ACCATGCCCT GCTTTGTCTA CTTTTTGAAT 360
 GATGTCYCAA ACTAGAAGST CTATTAAATTT AAAAAATTAA GSATAGCATG CCATAATTAA 420
 10 AAATAATAAC AGTGGGAAAA GGCACCTTTC AATGATTGAG ACATCAACTT GTGATTTAAA 480
 AAAACGAAAA ATAAATAATA GGAaaaaaag GCGAAAAACT TAAATAAAAA TAAAAATTAA 540
 15 AAAAAAAAAA AAAAACTCGA GGGGGGGCCC GTA 573

20 (2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 684 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CTCTTTGGCT GTGTCTACCT CCTTCATCTG CTCGGCCGAC ATAAGCACCG CCTGCCCCCT 60
 AGGCTCCAGC CCTCCCGCAC CAGCCCCAGG GCACCGAGAG CACGAGCATG GGCACCAAGC 120
 CAGGCCCTCC AGGTGTCTCT YCAGTCCCT TATGCCACTA TCAACACCAG CTCCTGCCCCA 180
 35 GCTACTTTGG ACACAGCTCA CCCCCATGGG GGGGCTCTCT GGTGGGGGTC ATTCCCGACC 240
 CAGGCTGCAC ACCGGCCCCA GGGGCTGGGC GGTGGGGCT CCACACCAT CCTGCACT 300
 40 GGCACCTTTG TCTCTGTTGA GAATGACTC TADGCTCAG CAGGGGAGAR GCTCTCTCAC 360
 ACTGGGCTCG GCTCACTCT TTCCCTGAC CTTGGGGGC CCAGGGCCAT GSAAGGATCC 420
 TTAGGASTTC GATGAGAGAG ACTAGCAGG CATTAGCTT TCCCTCTCC AGGCTCTCTG 480
 45 GGTGTCATCC CTTACTTTTA ATTCTTGGC CTGAATAAG TGTCCATAG GTGTGTGGC 540
 AGGCCACCT GTTCCGGATG TGGTCTGTGT GGTGTGTGG GCACAGGTCT GAGTGTGTA 600
 50 GTGACASTTA CCCCATTTC GTCAATTCCT GCTGAACCTA ACTCAGCAAC ACAGTTTCTC 660
 TCAAAAAAAAA AAAAAAAAAA AAAC 684

(i) SEQUENCE CHARACTERISTICS:

60 (A) LENGTH: 1036 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

	TGAGAGCAGA TGCACAGGAG AAAGTTCCCG CTCGCGACCG TCTCAGACCT GAGGCTGAGC	60
	TTGCACTGAG GGCCTCTCCT CGGCTCTCTG CCGGCTCCCA GAGCTGCCAT CCGTCTGT	120
10	ACAAGGCAGA GGAACCCGGA TGTGAGGCGC CAGATCACTT CCAGGGACTT GGGCTTCCCA	180
	TCTGAAATCC TTTATTTTTC TACCAAGGCG TGGCTCCCGG GTTGAGAAAG AAGAAGCACC	240
15	CTCTCCCGCG CCGTCTCTCT CTGACCCCTT GGGCTGTGTA CTTACTCTCT CCGCCAGGGG	300
	CGGGCGCGGG CCGCTGCGGA CCGCTTAAAG CCGAAGGTGG GCGCCAGGAC CTCTGCGCAG	360
	AGTGGAYTGC TCATGCGAGA TGTGTGCGAA TGTCTGGCTG WGTCTTTCCG GCAMCTGGCT	420
20	YCCCTTTCCG GGGTCCCTCT CCGTCACTGT GATCTGTCTT CTCTCTCCCG CCGTCACTAT	480
	GCTCTCTTGA GCGTTAGTCC AGGGGTCAC TCTCTCCACC CCACTCACTT CACAGGGTTG	540
25	TTCTGAGGCT GCACAGAGGA GCAAACTCCC TGAAGGCTCT CAGGCACTAT ATAGGGGCGG	600
	CGCACTTCCA GGTCTCCCTG GATGGGAAG ACCCAGCCCG AGGCTGCGG ATAACACTGT	660
	GTTTGCAAAAT GAGATTCAGT GATTGCGGTA TGCAGCTTGT GGGGAGCTGG CCGTGGCAGG	720
30	TAGGGGTAST TGGCTTGGCG TTCTCTTTTG TGATCCGAGC CCGAGGCACT TGCATTGCTG	780
	GCCCAGCCCG TGGCTTGGCG GCGGGGAGA GGCACAGAA GGGGCTGGCG AGGGGCGGTG	840
35	GAGGACTCAG GAATGCCCCG GCGAGACTCG GTATGGGCGT TGAGCCAGCG GCGCTCTCTG	900
	GTTTGACTTC CCGGATGCGT TCTTGCCTTC TCAGCTGTGT CCGACCCGAC CATGTAATAA	960
40	AACCCAAAGG AACAGCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1020
	CCCGGGGGGG GCGCCG	1036

45 (2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

50	(A) LENGTH: 908 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

	TTAAACAAAT GGAATCATGC AATATGTGAC CTTTTCGCTC TGGCTATATT TATTTAGCAT	60
	AATGTTTTTT AGGTTTCATC AAGCTGTAGC ATGTATCAGC ACCTCATTTT TTTTCTGGC	120
60	TGAATATTAT TCCATTATAT GATTTACCA CAATTCATTT ACCTATTCAT CTTTGTGTTT	180

341

TGCTGTCTGG CTATTGTGAA TAATGTTTCG ATAAATATTC ATATACAGST TTGTATGTGG 240
 CTTTATGTTT TCATTCTCTT TGGCTATGTA CATGGGAGTA GAATTCCTAGG TCATAATATA 300
 5 ATTTTATGTT TAACCTCTCA AAGAATTGCC AAAAGGTTTT TCATAGTGGC TGCATCATTT 360
 ACATTCGCAC CGGCAATGTA CAAGGATTTT TATTTTTTCA TATCCTTGCA CTTACCAACA 420
 10 CTTCTTTTTC GTWATWATTT TTTTTTTTCA TTATTCGCAC CCTAGTGGAT GTGAAATGGC 480
 ATCTTATGTT TTGATTTTGC AATTCTCTAA TGACAAATGA TATCTATCTT TTTTATGTC 540
 CTTACGGATC AAAGGTATTT CCTTGGAGAA ATGTCTCTTC AAGTCTCTTG CCATTTCAAA 600
 15 AATTGCTTAT TTGTCTTTTA TTATTCATTT TTAAGAAATT CTGGGCGAGC GCACTGGCTC 660
 ACCCTAATC MTAGCACTTT GGGAGGCCAA GGGGCGCAGA TCACTTGAAG TCAGGACTTC 720
 20 GAGACGAGC TGGGCAACAT GGTGAAAGCC CATCTTACTA AAAATACAAA AATTAGCTGG 780
 GCGTGTGGC AGGTGCATCT AATCTATCT ACTCAAGAG CTGAGGCGAG AGAATGGCTT 840
 GAACCCAGGA GCGGAGGCT GCACTGAGCC AAGATGAGC CATTGCACTC TAGCCTGGGT 900
 25 GACACAGA 903

30

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 655 base pairs
 35 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

40 TGCACTGGTT CCTTCTCCCC AGCAAATACT GCCTTCTTGT TTTTCTCTGA TGTGGCAGGT 60
 GACTACAAAA TCGGCTTGG TATTCTTCAA ATGCATATAF ATTCTTTTCT TCTGAGCTCC 120
 45 CTCTTTCTCT AGATTAGAAA ACTGCTTCAT TTCTCTCTCA CTGATCTTC ACTCCAGGT 180
 TGTCTCTCTC TCTTCCCCC CTGCTTCAGG TCTTCTTTT TTTTCTTTC CTGAGCAAT 240
 GGGCAGCAAA AGTTGTTTCA CAGTGAAAW TTAGGCATCC TCAAGTTTCT TCCAGCTTC 300
 50 TGCTGTGTTT TCTTAGAGTA AATTGCCAAT TTCTGTTTT ACAGGAAATC CTTTTTAAA 360
 AATGGAATCA CTGTGCTCCC CATCTACTCT GCAAAAATG CATTTTCTT TATTTTCAA 420

60

TCTTCTCTCT TCTTCTCTCT TCTTCTCTCT TCTTCTCTCT TCTTCTCTCT TCTTCTCTCT 480

AAAAAAAAA AAAAAAACY GRAGGGGGG CCGGTACCAA TTCGCCCTAT AATGA

655

5

(2) INFORMATION FOR SEQ ID NO: 89:

(1) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1102 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

20

TTTTTTTTTT ACCATTTAAA ATAAATGAA AGTGACCTTC TGTTTATAAA AATCTTTGTC 60

TGCATCTCTG CTTATTTCCCT TAGAAGAGAT TCCAAGAAGC GGTGAGTGAT TTCACGGGAG 120

CAGAGGGTTC GGACATATTA CCGGCGCGGA TCCCTCTTGG AGTGAGATGA CTCTCCGSAG 180

AGATTTAGTC GTCACTCTCG CGTGTGAGGC TGGGTACAC CCGAGGGATG TGTCTATCAA 240

25

GATCGAAGAT CTTTACACG CTCTGATTT TGTTCGCTT TTTTCTATT ACTAGTGAGA 300

AKGAAACTTT TTATATGATT ATTATGATC ATAATGCAAC ACAAATTACT GCTTCATGTT 360

CTTTACTTT CCGTGAAGG TTTTATGCG TTTTAAAAAT TGCTATATAT TAACTTTGTT 420

30

AATACTTCCA TGCTGATTT GTGGSCATCA RTTCCCGG GACAGGCNT GCATTTTTC 480

CCTTCACACG CTGGGTGGTT TTTATTTTC AMTCTATTT CTGTTCTTC TATGTTTTTA 540

35

TGTTGAGACG GCTTTCTCCG TTAGAAAGC AGTTTATGAA GATTTACTTT CGATAGTCTT 600

CTCTCTACTT TCTACAGTGA ATTCTCTGAT GTGTCTGGGA GTTTGGGGGT CTGGGTAAGA 660

RTCTCTCTCT CACCTATTTC TCTATTAGC TCCACAGCCT CATGCTTTAT GARATTGGTG 720

40

GCCGGGARGC GGGGAGATTT GCGATCCCC CAAGCCAGAC TTTATCCCCC TATCCCTGCC 780

TCTGGATCCC ACCTACAGGC CTGGGAATTC CTTGTGGTA GGGGCCAATG GTCTCGCACT 840

45

CTCACCTGTA CCGCAGGGCT GGCACAGSAT GGTGAAGGAG AGAGCTGCC CAAGCGCATC 900

CYTCTGGTGT CCCCCTGACA CGCTCCAAA GTGAGCAGGT AGGTTTCAAC AGCCCCAGGT 960

TGCAGGTGGG AGATGAAGCT CAGGCTGGAG ACCAGTATCT CACAGTTCTC TTTGCATGCC 1020

50

CGGTACTTGT TTAGTCAACT GATCAAGTGA AAATTCTAGC CCCAGAGGCA GGAGAATCCG 1080

GAACAAAATT AAACCAGCCA GG 1102

55

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 1533 base pairs

(E) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGCACGAGCC GNCACGGGCA GCGCCCCATA GGGCCAGGGA CCCCCTGACA GCGGGAGGCG	60
	CGGGTGAGG TTATGATCC AGCGGGCGGG CCGCGGGGGG TGCTCCGGG GCTTGGGGG	120
10	TGNTGCTGC TGCTGAACCC GGGCGGGGG AAGGGCAAGG CATTGACCT CTTCGGGAGT	180
	CAGGTCAGC CCGTTTTCGG TGAGGCTGAA ATCTCTTCA GCTGATGCT CACTGAGGG	240
15	CGGAACGAG CGGGGAGCT GGTGGGCTG GAGGAGCTGG GCGGTGGA CCGTCTGCTG	300
	GTCATGTYT GAGACGGGCT GATGACGAG GTGGTGAAG GCTTCATGG AGCGGCTGA	360
	CTGGGAGAGC GGCATGAGA AGCGGCTGCT TAGGCTGGA GAGGCTGCT GCAAGGCTCT	420
20	GGGAGCTTC TTAAACATT ATGCTGGCTA TAGGAGCTC ACCAATGAG AGCTCTGAG	480
	CAACTGACG CTATTGCTCT GCGCGGGGCT GCTGTGACCC AGGAATTCG TGTCTGCTA	540
25	CAAGGCTTCG GGGCTGCGGC TCTTCTCTCT GCTGAGGCTG GCGTGGGGCT TCATTGCTGA	600
	TGTGACCTA GAGAGTGAGA AGTATGGGG TCGGGGGGAG ATGGGCTTCA CTCTGGGCA	660
	CTTCTGCTCT CTGGGAGGCC TGGGACCTA CCGGGGGGGA CTGGGCTAGC TGGCTTAGG	720
30	AAGATGGGCT TCGAAGACAC CTGCTTCCCG GGTGTGGCTC CAGGAGGGCC CGTAGATGC	780
	ACACTTGTG CCACTGGAGG AGCGAGTGG CTCTCACTGG ACAGTGGTGG CGGACGAGGA	840
35	CTTTGTGCTA GTCTTGGGAC TGCTGACCTG GCAGCTGGGT AGTGAGATGT TGGTGCAGC	900
	CATGGGCGCG TGTGAGCTG GGTGATGCA TGTCTTCTAC GTGGGGGGGG GAGTGTCTCG	960
	TGGCATGCTG CTGGGCTCT TCTTGGGAT GGAGAAAGGG AGGCATATGG AGTATGAATG	1020
40	CGCTACTTGG GTATATGTGC CGGTGGTGGC CTTCGGCTTG GAGGCGAAGG ATGGGAAAGG	1080
	TGTGTGTCGA GTGGATGGGG AATTGATGCT TAGGAGGGCC GTGGAGGGCT AGTGCAGCC	1140
45	AAAGTACTTC TGATGCTCA GCGGTGGCTG GAGGCGGGGG CGGAGCTGGA AGCTTGAGCA	1200
	GATGGCAGCG CGAGAAGAGT CATTATGAG CCGGGGGGGG GTGTGCTCTT AGTGTCTACT	1260
	TGCAGGAGCC TTCTCTCTTC CTTAGGCTG CAGGGGCTGT CCACAGCTTC TGTGGGGGTG	1320
50	GAGGAGACTC CTCTGGAGAA GGTGAGAAAG GTGGAGGCTA TGCTTGGGG GAGAGGGCA	1380
	GAATGAAGTC CTGGGTGAGG AGCGGAGCTG GCTGGGCGCA GCTGGCTATG TAAGGCTTTC	1440

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

ATCCTCTGGA ATCTAGGTGS AAGCCACCAA GCCTTCTTCA CACTTGCGTT CTGAGCATCT 60
 GCAGACTTAA CCCCATGTGS CAATCAGCAA GGCTTATGGC TTGTGTCTTC CAGAACTGTG 120
 15 GCCAGAGCTG TAACCTGGCC CCTTTGAGCT GAGGCTGAAG CCAGAGTCTG AAGCTCAGCA 180
 GGGCAGTARG GCGCTGGGCG TGGCCCTTGA AACCATTCCT TTCTCCTAAG CCTCTGGGCC 240
 20 TTGATGGGA RGGGCTGTCC TCAAGATTTT TGAAATGCCT TTGGAGGTTT TTTGCGTTGT 300
 CTGATATATT GGCTTCCTTT TAGTTATGCT CATCTCTCTA GCAAGTGAAT GTTTCACAAC 360
 CTGCTTGGAT TCTTTCTCTA CCACAGAGCC AGGCTGCAAA TTTTACAAAC TTTTAACTTC 420
 25 TTTTTCCCTT TTAAATATAA ATTTCAATGT TAAGTCACTT CTTTGCTCCC ATATCTGATT 480
 TAGGTTGCTG GAAGTAGCCA AGTCACCTCT TGAATGCTTT GCTGCTTAGA AATTTCTCTT 540
 30 ACTAGGTAGC CTGGGTCATC AACTTAAAT TCAAA 575

35 (2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 639 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

45 TCCTTTTCATC TTAAGCACCA CCGACAGGG CAGGTACTAT TACCATCTCC GTTTCACAGA 60
 TNAGGAACCT GGCACAGGAA GCATTTAAGT GGATTCCCCA GGATCGCCCC ACTGTCAGGA 120
 GCAGANTCAG AATGGGCGTC AGCATCAGGC TCCCAATCCT GGCTTTTAAC TGCTGCGCTC 180
 50 TGCCCTTCYC TCWCCCCACC TCCCCACTCC AGTGCCCTTG GTCATGCCAC TGCAGCTTTC 240
 AGGCCAATAC TGGATTAGCC TCTTAGTGT CTGTGCCCTG CAGCCATTTT CCGAGGCAGC 300
 55 AATTCATGT GCCCTCACTG ATGTAGGTGG CTCTTGTGTC ATTTGTCCAC TCCATATTGAA 360
 TTGTTTATGC ATCTTGTTC CACTCAGAGC ACCCTCCCTC TCACAGTCC TCCTTATAAA 420
 AATGTCCCTC AGTGTCTGCT ATGAGCCAGG TGCAGACTTA AGTGACAGG CTGCTACGGG 480
 60

345

AAATAAAAAA TTAACAAGGA GCACCTGGCT CTTAATGCAC ACTAACAAAC TATGTTAAGT 540
 GTCAGGAAGG AAGGTTAAG SATGCCAGGA AGGCTTTTAA TAAATAAGCT GACTTAGATG 600
 5 GGCAGGTGGT GCTGARGATT AAGAACGTGT TCTTCTCGA 639

10 (2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 744 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

20 GAATTGGGCA CGAGACTGGC TGGAGTCTGG CTGCAGAGGG AAGACATCAG CAGGGAGGGA 60
 GCGAGGGGCT CTCACATCTT TCTCTGGGC APTGTCTCTG TCTTTGTAAG CCCAGAATCT 120
 CCCCCTCTCT GAAGGGAGGC CAGCACCCTA GAGGGGCAGC AGGTCTGCTG TGAGGGTTGG 180
 25 AGTAGTGTGA GAGCTCAGCG TACTACTAGAA TCGCCATGSA CACCATGTGG GGGTCTCTG 240
 GGTGAGGCA CAGAACAGTG TCTTCTCTGC TCTCTCTCC CTGCAGCTTC CCCCACCTT 300
 30 GTNGTTTATT TGGTTTGATA CCAATCAGCA GACCTTGCAA GGTGGAAGCT CCCAGGCTCT 360
 CAGTCCAGG ACTCTCATGT GCCAGTCAIC INTACTGTAA CTGCCCAATG AGTACTTCTT 420
 GGGCACTGGC AAGATAGAGC CAGTTTACCA AGACAGGGGA APTCAGTAG AGAAACAGTT 480
 35 GAATATACAT AGACCTAGCT AATGGGAGA GTGGAGTTTT CTATTACTT AAATCAGCT 540
 CCGTAAAAAT TCAGAGTGA GAATTTTTCA AGACAGTTT GTGGGSCAGG CCTAGGGAAT 600
 40 GGATCTCTCT GATTGGCTAG GGATGCAATC ATAGGGGTGT AGAAAAGTWC CTTGTGCACT 660
 GATTCACATT CTGCTAGAG CTACCAAGSA GTTGTCTGTC TGTGCTGCC GTTAGAGCA 720
 TCTGCTGCA GAATGAAA AGTG 744

50 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 526 base pairs

(B) TYPE: nucleic acid

60 GAAAGGAAAT TGTCTAGG AGGCTTTCA AGAAGCTT TGGATGAT CAGAGACA

346

AAGGCCATAA GTGCTGGCCT GTTGGGACAA ATGAGACAAA TCCCATAGGG TGSTGATGAC 120
 AGGSCAYTCA GGCATCTAY TCTGGGGAA AATGAACTT GTGCTCCTAT CAAATGCTCA 130
 5 GTTGTA AAC TG3AAAAAA TTTAGAAGA CATCTGTCC AGCATCTGTG TTTATGTCTA 240
 TAAATGTAG AAACTAAAG CACAGAGATG TTAATTTTT TGTCCAAGGT CCAACAGCTG 300
 10 GTTAGCARGC TTGCTCTGCT GACCTTTCTA CTGAACACAA GTGCGCTGG GGSAACTCCT 350
 CAGACAGAT GGTGCTGCT ATAGCTGGG TATGGGCACT ATTAGTAGTT AACCACTCAA 420
 CCAATTTCC CATAGTCTAG GTCCTGCTTC AGCTGAGGT TAGGGAAAAA CACAA3AAAA 480
 15 TCTTTACCA CTCTACCACT GCTGGGGAT GACTAAGAG ATCCCC 526

20 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

30 GGCATAGGGC AGGAGAGACT TGCTCATGG GGAGAAGCCT CCAGTATAGA TGGGACCTCC 60
 AGGAGCCCAA GTAGCATAGA CCTGCTGAT CCGGGCCAT TGAGCCAGAG GATTGGGCT 120
 35 GAATGTCCCC AGAGACAAA GGSAAAGTA GATCTTTTC CTAAAGATG AAAGCCATCG 180
 CCGGGCTTG CTATATGCTC TCTCTCTGG TCCTCCACA TGTGTTTCT GAACATTTGT 240
 TCTGCATCA CAATCCCGT CATCTGTCA TCTGGCCCTT CCCACCTTC CACTTATCT 300
 40 CTGCACTGT CTCGCTGTC ACCTGGCACC TGGGTGAAG CTGCTCTT CTGCTGCCA 360
 TAGCTCCAG TGTATGCTT TGACTCCCC AGCCATATG ARACCCACT CAGGAGGGCC 420
 45 CCTGGA 426

50 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 844 base pairs
 (B) TYPE: nucleic acid
 55 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60 GGCACAGCG CACGAGATAG GAAGCTTGGC AGGGGCAGCT CCCCCAGTGC GCATTGCCCT 60

5 GTAACTCGAG CGCCTGGGAG TGGGAGAGAG CTTGGAAATG GAGTAGGATG GTGGACCTCG 120
 TCTTCTGCTG CTEATCCAG GCTCTCTCA TAACACCTAC CTAGCAGGCG CTGGGGACTT 180
 CCGAGCCCAA GGAACAACCTG AGAATACTGA GTGCCAGGCT AGCCTAGCC CCATTTTACA 240
 CCTGGGCAAA GTGAGGTGAC TGGATTCAAA CATTAGATT TAAACCTCCT CTGTGTCTGT 300
 AGCACCTGTA TATAACTGCC AGCCTCTGCT GCGCCTCTCC AAAAAGTCTC TGCCTTGTG 360
 10 TTTGGCAGCT GTCTCTGTG TCCCATTTCT CTGCTCCTCC TTTCTCAAC TCAGANTCAG 420
 CCTGTTAGTT CAGCAAATGT TCATCGAGCT CCATAATGTA GCAGGACAGG NCTGTCTAAC 480
 15 AGATTCTGGN CTTGCAAGCG TGAGACAAGT ACTCTCCATC TTTCTCTCAT CTTACAGAT 540
 GGTCTGCTCA ACAACTTTGC ACTGAATTGT AAATAATTGA TATTGCATAA AACATTGATG 600
 TTCTTTAAGG GTAGCCAGC AAGTTGGCAA GTCTTATGAT GATAACTGCT CAAGCATCTC 660
 20 TCAGTGAAGC ATTTGGTGT OCTAGTCTG CCTATGCTG AGTCAGCTA TCTACGCGCA 720
 TCTACTTCCA CNTGCCGCC CATGCCAGGC TCACCTTAG CTAGATGCC TGAGCAGGTC 780
 25 GCASAAAGGA GGCACCTGCT TTATGCTTGG GAGCCACAAA CTCTCTATC CAGAGACAG 840
 TTTT 844

30

(2) INFORMATION FOR SEQ ID NO: 97:

35 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1935 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

AGCCCTGCTG AAGTATAGGT TCTTCTATCA GTTTCTTTTG GGCATGAAC GAGCAACAGC 60
 AAGGAGATT AGGATGAAAT ATCTGAGAGC GTGAGTAAG ATTACTGTCT CTACTAGCG 120
 45 CTCTTACCTG GCGGGCTCA TGAGGTGCA GTATGAAGAA CTGGTCAAA AAGATGATCT 180
 AATGGTGTG CAAGATACAG CAAGAAAGG ATTCTYCTCA AAGCCATGCC TCGGCAGCAG 240
 50 GAACACCATT TTCACCTAG GAACCGCGCG CTCTGTGATC TCCCCACTG AATTGAGGC 300
 CCCCATCCTG GTCCCTCACA CAGCGAGCG GAGAGCAGA GATATGATTT TGAGGCGCTC 360

60 ATATTAATA TACCTTAAA AATTTGAT CTATATTA GATATTA CA TATTTAT 384
 CTCTTTTCTG TGGTATCA CATTCTTTCT CTCTTATCA AATTTGAT AATTTGAT 400

	GTTCCTGCCG TGGACAGGTA CTGGGAACA GGTGCTTGGC TTGCTATGBC CACGCTTTGA	660
5	ACTGATCTCTG GAGATGAATG TTGAGAGGCT GCGAAGCACT GACGCCAGG GCTAGGGGG	720
	GTTGGATACT GGGGCCCACT ATATCAGAGG CCGCTATGGA GAGTTCTGCT CCGCTCTCT	780
	CAGTATCAAC CAGACAATTC CTAATGAAAG GAGCATGCAA TTGCTGGGAC AGCTGCAGGT	840
10	GGAGCTGGAG AATTTTGTCC TCGAGTGGC AGCTGAGTTC TCTCAAGGA AGGAGCAGT	900
	TGTCTTTCTG ATCAACAACCT ATGACATGAT GTTGGCTGTG CTGATGGAGC GGGCTGCAGA	960
15	TCAGAGCAAA GAGGTTGAGA GCTTCAGGCA GCTGCTCAAT GCTGGGACAC AGGAATTCAT	1020
	TGAAGATTG CTCTCTGCCG CTTTGGGGG TTTAGTGGCA TTTGTGAAGG AGGCTGAGGT	1080
	TTGATTTGAG CTGGACAGG CTGAGGCACT TCGAGGGGAA GAAGGCGGG TAACTTAGCT	1140
20	GATTCCTGGC TTGCTAGTT CTTGGAATC ATCAGTGGAA TCTCTGAGTC AGGATGTAAT	1200
	GGGAGTTTC ACCAACTTCA GAAATGGCAC CAGTATCATT CAGGAGGCG TACGCCAGCT	1260
25	GATCAGCTC TATCATGGCT TCGACGGGT GGTCTGCCAG CCGCAGCTCC GAGGCTTCC	1320
	TGCGGGGCT GAGTCACTA ACATTCAACA CTTTATGCTG GAGTCAAGA AGCATAAGCT	1380
	CAACTTCTGA TGGGACAGAA ACCGCTCTGA GATCTGCGG TCATCTCCAT GGAATTCTGC	1440
30	ACGCTATTC ATACGCTTCT TCACCTGGGG TACCGCTTCC AGTTTTCGCC TTGCTTCCCA	1500
	GGGCTTGAAC ATGGCTTACC TGCTTCACT CCGAGCAGCT TGCCCAACAG GATAAGCTGG	1560
35	ATCGGCTTTC CTTCTGAAT ATCCAGTCT CTTCAAGTTT CCGAAGACCA CTTGCTTGTG	1620
	GGCTTCCAAA ATGGCTTCA TCAATTTCTC AGTCTGTGAC CTTCTTTTCC TCTTCCATA	1680
	CACGCAAGAG TGTCTTCTC CCGTGAAAA ACCAATGCGT CAATCTCTGG TTCACTCAAC	1740
40	TAGTCACCAT GTCTGAGGC ATGAAGCTC CTCAGCTCTT GGAATTGCTG GCAAGGGGTG	1800
	ACTGCTCTG AGTCATTGTG TTTTCAAAG TGATTTCTTT TCTGTAGCTT TTTGACCTAA	1860
45	GATCTCAGCA ATTGAACAC TAACCTCTCC CTTCTGGCT CAAGAATTAC TCCGAAGTCA	1920
	GTCTGCAGAA AATAAATATT TAGTATGACA TGAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1980
	AAAAA	1985
50		

(2) INFORMATION FOR SEQ ID NO: 98:

- 55 (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1416 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

	ATATGAAGGG AAAGAATTG ATATGTTTT CTAAATTGAT GTCAATGAAG GTGGACCATC	60
5	ATATAAATTG CCATATAATA CCAGTGATGA CCTTGGTTA ACTGCATAGA ACTTCTTACA	120
	GAAGAATGAT TTGAATCCTA TGTTCCTGSA TCAAGTAGCT AAATTTATTA TTGATAACAC	180
10	AAAAGGTCAA ATGTTGGGAC TTGGGAATCC CAGCTTTTCA GATCCATTTA CAGGTGGTGG	240
	TCGGTATGTT CCGGGCTCTT CGGGATCTTC TAACACACTA CCCACAGCAG ATCCTTTTAC	300
	AGGTGCTGGT CGTTATGTAC CAGGTTCTGC AAGTATGGGA ACTACATCG CCGGAGTTGA	360
15	TCCATTTACA GGAATAGTG CCTACCGATC AGCTGCATCT AAAACAATGA ATATTTATTT	420
	CCCTAAAAAA GAGGCTGTCA CATTGACCA AGCAAAACCT ACACAAATAT TAGGTAAACT	480
20	CAAGCAACTT AATGAACTG CACCTGAAAG GAAGAAGTTA ACTGAGGATG ACTTGATACT	540
	TCTTGAGAAG ATACTGTCTC TAATATGTAA TAGTTCTTCA GAAAACCCA CAGTCCAGCA	600
	ACTTCAGATT TTGAGAAAG CTATTAACCTG TCCTGAAGAT ATTGCTTTTC CTGCACCTGA	660
25	CATTCTTCGG TTGTCAATTA AACACCCGAG TGTGAATGAG AACTTCTGCA ATGAAAAGGA	720
	AGGGGCTCAG TTCAGCAGTC ATCTTATCAA TCTTCTGAAC CCTAAAGGAA AGCCAGCAAA	780
30	CCAGCTGCTT GCTCTCAGGA CTTTTTGCAA TTGTTTTGTT GGGCAGGAG GACAAAAACT	840
	CATGATGTCC CAGAGGGAAT CACTGATGTC CCATGCAATA GAACTGAAAT CAGGGAGCAA	900
	TAAGAACATT CACATTGCTC TGCTACATT GGCCCTGAAC TATCTGTATT GTTTCATAA	960
35	AGACCATAC ATTCAAGGGA AAGCCCAATG TTGTCACTA ATTAGCACAA TCTTGAAGT	1020
	AGTACAAGAC CTAGAAGCCA CTTTTAGACT TCTGTGGCT CTTGGAACAC TTATCAGTGA	1080
40	TAATTCAAAAT GTGTACAAT TAGCCAAATC TTAGGTGTT GATTCTCAA TAAAAAGTA	1140
	TTCTCAGTA TTAGAACCAG CTAAAGTAAG TGAATGCTGT AGATTATCC TAAATTGCT	1200
	GTAGCAGTGG GGAAGAGGGA CGGATATTTT TAATTGATTA GTGATTTTTT CTTACATTT	1260
45	GCATGACTG ATAACAGATA ATTAAAAAAA GAGAATACGG TCAATTAAAT AAAATTTTAC	1320
	ATCTTGTAAG GTGCTGGGA GGGGAACAG AAATAAAATT TTTGCACTGC TGAATAAAAA	1380
50	AAAAAAAAAA AAAAGGAAAC TCGAGGGGGG CCCCCG	1416

(1) TYPE: nucleic acid
 (2) STRANDEDNESS: double
 (3) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5	NTCTAGCCCTA ATCAAGATGG GGACATACTT CGGAGGAGG TTCTTCATGA ACATATCCAG	60
	AGATTGCTCTA AAGTAGTGAC TGCAAATCAC AGAGCTCTTC AGATACCAGA GCTTTATCTT	120
	CGAGAAGCAC CATGGGCATC TGCACAATCA GAAATGAGGA CAATAAGTGC TTATAAAACC	180
10	CCCCGGGACA AAGTGCAGTG CATCCTGAGA APTGCTCTTA CGATTATGAA CTTCTCTGAGC	240
	CTGGGCAATG AGGACTCTGT CCTTGGAGCG GATGACTTTC TTCCTCTGTT GCTCTTTGTG	300
15	TTTATAAAAG CAAATCCACC CTCTTTCTCT TCTACTCTGC AGTATATCAG TAGCTTTTAT	360
	GCTAGCTGTC TGTCTGGAGA GGACTCTTAT TGTGATGC AGTTCCACAG AGCAGTAGAA	420
	TTTATTAATA CCATGATGA CGGAAAGTGA CCAAGACCAA GGGCCACCAA GGCAGCAGAC	480
20	TGTTAATCAG ACAAAACAGT CTCTGAGAAG GTGATCAGC TGTCTTGAAG GCTGAAGATT	540
	GTTTCTGATG ATACTGCACA GCATCAGGCA TTTTAAAGCA GATCTTTACT AAACAGCTTA	600
25	ATGAGCTAAC AAGCAAGTTC TCTCTCTTTT GGGCTCTTTC CTTTCTGAGT TGCATATTCT	660
	ATTTCTTTCT CCCCAGTAG AGACTAGTAC TACAAAAAGG GACCACATTT TTCAAATATT	720
	TCTAAGTATA AAAAACAAAA CAAAAATCTC TTAGGAAATG TCTAGACCTC CATTCTTGGA	780
30	TTCCCTTTCT TCCCTTTTAT TTTAAAAAAG AACAGTACCC CTCTTTTAAG ATGCTCTCTT	840
	ACATTAATGA GATCTAATG GAAAGAAGGT ATGATTTGCA CTGAGGATTA GAATAGTGGT	900
35	GGCTTAGTGG CATTATCTAT AAATACACTC ACCTAAATTC AAAGCTAAGA AGGAAATGTA	960
	AATATAATAT ATATTTATAT TTGATSTAAT ATGGACATCT GGAGATCTTA ATAAACAAGG	1020
	ACTATCTCTG ATAGTAGGCT GTGACATACT GTCTTCTGAA ATGGTTTCTT TGACAAAAATT	1080
40	TAAGCTGAGC TTAAAAAGCA AAAAACAAAA AGTACACAGA AATATTTATT AAAATGTAAT	1140
	ACAGTTTATT GAACCTTTCTA GGTATGGAGT TTGATGGACA GGGCTGCTT TAATGAGTGT	1200
45	GAAGGTCACT AAGTCACCTA GACATCTCAC GGTGGAAGTT TGTGAGCTTG CATTAGGAGA	1260
	TAGACTGATT ACCATACATG ACATAAAAAAG GAACAGTGGG TAGCTCATAC TTTATGGTGG	1320
	TTCTCTCTCT CGGAAATAAT ATACTGCAGA AATCCAGAC AGAGCTCTTT ACAAACTTTT	1380
50	AATCTTAATA TATTTTGTAT GATTATTCAC ATTGAATGCA CAGACCAAGA ATTCAAGTGAA	1440
	TCTCATTTTT TAAAAACTA ATTTGTATTG TCTCTCTAG TGATACAAGT TTTACTAGTG	1500
55	ATAAACTATT TTAATCAACC ATACTATTCT TATGAAAAA AATATCTATT TTGGCAGGTT	1560
	TCTGTGCTTT TATTTCCCTC TTCTGAAAAA AAGTCTGTGT TTTCATAGTT TGGTTTGCAT	1620
	TGTATATCAA TAATTAATCA GGAATGGGTT TTGTGCTCTG AAAAATTGGC CATGGAGGCA	1680
60	CACCAAAAGT TCAAGCACAA GTCTGTGACA TGAGGCATCA CTCTCTGGTT TCACTTCGTG	1740

10
 5
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000
 1005
 1010
 1015
 1020
 1025
 1030
 1035
 1040
 1045
 1050
 1055
 1060
 1065
 1070
 1075
 1080
 1085
 1090
 1095
 1100
 1105
 1110
 1115
 1120
 1125
 1130
 1135
 1140
 1145
 1150
 1155
 1160
 1165
 1170
 1175
 1180
 1185
 1190
 1195
 1200
 1205
 1210
 1215
 1220
 1225
 1230
 1235
 1240
 1245
 1250
 1255
 1260
 1265
 1270
 1275
 1280
 1285
 1290
 1295
 1300
 1305
 1310
 1315
 1320
 1325
 1330
 1335
 1340
 1345
 1350
 1355
 1360
 1365
 1370
 1375
 1380
 1385
 1390
 1395
 1400
 1405
 1410
 1415
 1420
 1425
 1430
 1435
 1440
 1445
 1450
 1455
 1460
 1465
 1470
 1475
 1480
 1485
 1490
 1495
 1500
 1505
 1510
 1515
 1520
 1525
 1530
 1535
 1540
 1545
 1550
 1555
 1560
 1565
 1570
 1575
 1580
 1585
 1590
 1595
 1600
 1605
 1610
 1615
 1620
 1625
 1630
 1635
 1640
 1645
 1650
 1655
 1660
 1665
 1670
 1675
 1680
 1685
 1690
 1695
 1700
 1705
 1710
 1715
 1720
 1725
 1730
 1735
 1740
 1745
 1750
 1755
 1760
 1765
 1770
 1775
 1780
 1785
 1790
 1795
 1800
 1805
 1810
 1815
 1820
 1825
 1830
 1835
 1840
 1845
 1850
 1855
 1860
 1865
 1870
 1875
 1880
 1885
 1890
 1895
 1900
 1905
 1910
 1915
 1920
 1925
 1930
 1935
 1940
 1945
 1950
 1955
 1960
 1965
 1970
 1975
 1980
 1985
 1990
 1995
 2000
 2005
 2010
 2015
 2020
 2025
 2030
 2035
 2040
 2045
 2050
 2055
 2060
 2065
 2070
 2075
 2080
 2085
 2090
 2095
 2100
 2105
 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250
 2255
 2260
 2265
 2270
 2275
 2280
 2285
 2290
 2295
 2300
 2305
 2310
 2315
 2320
 2325
 2330
 2335
 2340
 2345
 2350
 2355
 2360
 2365
 2370
 2375
 2380
 2385
 2390
 2395
 2400
 2405
 2410
 2415
 2420
 2425
 2430
 2435
 2440
 2445
 2450
 2455
 2460
 2465
 2470
 2475
 2480
 2485
 2490
 2495
 2500
 2505
 2510
 2515
 2520
 2525
 2530
 2535
 2540
 2545
 2550
 2555
 2560
 2565
 2570
 2575
 2580
 2585
 2590
 2595
 2600
 2605
 2610
 2615
 2620
 2625
 2630
 2635
 2640
 2645
 2650
 2655
 2660
 2665
 2670
 2675
 2680
 2685
 2690
 2695
 2700
 2705
 2710
 2715
 2720
 2725
 2730
 2735
 2740
 2745
 2750
 2755
 2760
 2765
 2770
 2775
 2780
 2785
 2790
 2795
 2800
 2805
 2810
 2815
 2820
 2825
 2830
 2835
 2840
 2845
 2850
 2855
 2860
 2865
 2870
 2875
 2880
 2885
 2890
 2895
 2900
 2905
 2910
 2915
 2920
 2925
 2930
 2935
 2940
 2945
 2950
 2955
 2960
 2965
 2970
 2975
 2980
 2985
 2990
 2995
 3000
 3005
 3010
 3015
 3020
 3025
 3030
 3035
 3040
 3045
 3050
 3055
 3060
 3065
 3070
 3075
 3080
 3085
 3090
 3095
 3100
 3105
 3110
 3115
 3120
 3125
 3130
 3135
 3140
 3145
 3150
 3155
 3160
 3165
 3170
 3175
 3180
 3185
 3190
 3195
 3200
 3205
 3210
 3215
 3220
 3225
 3230
 3235
 3240
 3245
 3250
 3255
 3260
 3265
 3270
 3275
 3280
 3285
 3290
 3295
 3300
 3305
 3310
 3315
 3320
 3325
 3330
 3335
 3340
 3345
 3350
 3355
 3360
 3365
 3370
 3375
 3380
 3385
 3390
 3395
 3400
 3405
 3410
 3415
 3420
 3425
 3430
 3435
 3440
 3445
 3450
 3455
 3460
 3465
 3470
 3475
 3480
 3485
 3490
 3495
 3500
 3505
 3510
 3515
 3520
 3525
 3530
 3535
 3540
 3545
 3550
 3555
 3560
 3565
 3570
 3575
 3580
 3585
 3590
 3595
 3600
 3605
 3610
 3615
 3620
 3625
 3630
 3635
 3640
 3645
 3650
 3655
 3660
 3665
 3670
 3675
 3680
 3685
 3690
 3695
 3700
 3705
 3710
 3715
 3720
 3725
 3730
 3735
 3740
 3745
 3750
 3755
 3760
 3765
 3770
 3775
 3780
 3785
 3790
 3795
 3800
 3805
 3810
 3815
 3820
 3825
 3830
 3835
 3840
 3845
 3850
 3855
 3860
 3865
 3870
 3875
 3880
 3885
 3890
 3895
 3900
 3905
 3910
 3915
 3920
 3925
 3930
 3935
 3940
 3945
 3950
 3955
 3960
 3965
 3970
 3975
 3980
 3985
 3990
 3995
 4000
 4005
 4010
 4015
 4020
 4025
 4030
 4035
 4040
 4045
 4050
 4055
 4060
 4065
 4070
 4075
 4080
 4085
 4090
 4095
 4100
 4105
 4110
 4115
 4120
 4125
 4130
 4135
 4140
 4145
 4150
 4155
 4160
 4165
 4170
 4175
 4180
 4185
 4190
 4195
 4200
 4205
 4210
 4215
 4220
 4225
 4230
 4235
 4240
 4245
 4250
 4255
 4260
 4265
 4270
 4275
 4280
 4285
 4290
 4295
 4300
 4305
 4310
 4315
 4320
 4325
 4330
 4335
 4340
 4345
 4350
 4355
 4360
 4365
 4370
 4375
 4380
 4385
 4390
 4395
 4400
 4405
 4410
 4415
 4420
 4425
 4430
 4435
 4440
 4445
 4450
 4455
 4460
 4465
 4470
 4475
 4480
 4485
 4490
 4495
 4500
 4505
 4510
 4515
 4520
 4525
 4530
 4535
 4540
 4545
 4550
 4555
 4560
 4565
 4570
 4575
 4580
 4585
 4590
 4595
 4600
 4605
 4610
 4615
 4620
 4625
 4630
 4635
 4640
 4645
 4650
 4655
 4660
 4665
 4670
 4675
 4680
 4685
 4690
 4695
 4700
 4705
 4710
 4715
 4720
 4725
 4730
 4735
 4740
 4745
 4750
 4755
 4760
 4765
 4770
 4775
 4780
 4785
 4790
 4795
 4800
 4805
 4810
 4815
 4820
 4825
 4830
 4835
 4840
 4845
 4850
 4855
 4860
 4865
 4870
 4875
 4880
 4885
 4890
 4895
 4900
 4905
 4910
 4915
 4920
 4925
 4930
 4935
 4940
 4945
 4950
 4955
 4960
 4965
 4970
 4975
 4980
 4985
 4990
 4995
 5000
 5005
 5010
 5015
 5020
 5025
 5030
 5035
 5040
 5045
 5050
 5055
 5060
 5065
 5070
 5075
 5080
 5085
 5090
 5095
 5100
 5105
 5110
 5115
 5120
 5125
 5130
 5135
 5140
 5145
 5150
 5155
 5160
 5165
 5170
 5175
 5180
 5185
 5190
 5195
 5200
 5205
 5210
 5215
 5220
 5225
 5230
 5235
 5240
 5245
 5250
 5255
 5260
 5265
 5270
 5275
 5280
 5285
 5290
 5295
 5300
 5305
 5310
 5315
 5320
 5325
 5330
 5335
 5340
 5345
 5350
 5355
 5360
 5365
 5370
 5375
 5380
 5385
 5390
 5395
 5400
 5405
 5410
 5415
 5420
 5425
 5430
 5435
 5440
 5445
 5450
 5455
 5460
 5465
 5470
 5475
 5480
 5485
 5490
 5495
 5500
 5505
 5510
 5515
 5520
 5525
 5530
 5535
 5540
 5545
 5550
 5555
 5560
 5565
 5570
 5575
 5580
 5585
 5590
 5595
 5600
 5605
 5610
 5615
 5620
 5625
 5630
 5635
 5640
 5645
 5650
 5655
 5660
 5665
 5670
 5675
 5680
 5685
 5690
 5695
 5700
 5705
 5710
 5715
 5720
 5725
 5730
 5735
 5740
 5745
 5750
 5755
 5760
 5765
 5770
 5775
 5780
 5785
 5790
 5795
 5800
 5805
 5810
 5815
 5820
 5825
 5830
 5835
 5840
 5845
 5850
 5855
 5860
 5865
 5870
 5875
 5880
 5885
 5890
 5895
 5900
 5905
 5910
 5915
 5920
 5925
 5930
 5935
 5940
 5945
 5950
 5955
 5960
 5965
 5970
 5975
 5980
 5985
 5990
 5995
 6000
 6005
 6010
 6015
 6020
 6025
 6030
 6035
 6040
 6045
 6050
 6055
 6060
 6065
 6070
 6075
 6080
 6085
 6090
 6095
 6100
 6105
 6110
 6115
 6120
 6125
 6130
 6135
 6140
 6145
 6150
 6155
 6160
 6165
 6170
 6175
 6180
 6185
 6190
 6195
 6200
 6205
 6210
 6215
 6220
 6225
 6230
 6235
 6240
 6245
 6250
 6255
 6260
 6265
 6270
 6275
 6280
 6285
 6290
 6295
 6300
 6305
 6310
 6315
 6320
 6325
 6330
 6335
 6340
 6345
 6350
 6355
 6360
 6365
 6370
 6375
 6380
 6385
 6390
 6395
 6400
 6405
 6410
 6415
 6420
 6425
 6430
 6435
 6440
 6445
 6450
 6455
 6460
 6465
 6470
 6475
 6480
 6485
 6490
 6495
 6500
 6505
 6510
 6515
 6520
 6525
 6530
 6535
 6540
 6545
 6550
 6555
 6560
 6565
 6570
 6575
 6580
 6585
 6590
 6595
 6600
 6605
 6610
 6615
 6620
 6625
 6630
 6635
 6640
 6645
 6650
 6655
 6660
 6665
 6670
 6675
 6680
 6685
 6690
 6695
 6700
 6705
 6710
 6715
 6720
 6725
 6730
 6735
 6740
 6745
 6750
 6755
 6760
 6765
 6770
 6775
 6780
 6785
 6790
 6795
 6800
 6805
 6810
 6815
 6820
 6825
 6830
 6835
 6840
 6845
 6850
 6855
 6860
 6865
 6870
 6875
 6880
 6885
 6890
 6895
 6900
 6905
 6910
 6915
 6920

CACTTTGAC ATCCACTTGT CACCAAATCK RGTTCATTCT GCATCCTAAG TAAGTCCTTT 240
 GATTGCTCCA GTTGTTCATT AGTAATGTCT CAARNGTAAT TTTTCTAGT AGTTTTCAGC 300
 5 CTGTCTTTTC KGGCTTCACT CTAACTTCT CCAGTACATA KGGCACATG TTGTGAGCAK 360
 GATCAWATTT TATTTAAAAA TACTTTACAW AKSTTTATEK CCAAATATTA GRAAATACAG 420
 10 ATTCATGGAA AGAAAAATCA CTCTCCCAAG GAGGTCACTG GCATGGTGAG GTTAAGGGGT 480
 GATTTTAATT TTTAAAAATG TATATTTTTT CTTGTGTAGA CTAGTAACAC CTTTGAAAAC 540
 ACAWTCCCTT GTAAAGTCTC TAATTTTSTA CTCCGATCT AGSTGRTCT TTCTTTCTCA 600
 15 GATATTTTAC AATTTCAITTT ATCAACCACT TTCTCTAGCC TTACCCCTC TCTTCAATAT 660
 TWACATATGC AGAAGTTTCT CTTAACAAAC ACCTGCTCT CCCTAGTTC TCTACCACC 720
 20 CTGTGCTTT CTTCCTCTC ACAATCAAAT TTAAGACTCT CAAAAAATA AAAAAAATC 780
 TCGA 784

25

(2) INFORMATION FOR SEQ ID NO: 102:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1035 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

AGAGGCCTG CTGCGTTGCT CTATCTCGT CTCCGCCAC CACTTAGCCT TTAGGCATC 60
 AATTACCAGT AGTTCTCCG CCACTATCTG GAAAATTACC CGATTGCTC CGGCAGAATA 120
 40 CAAGAGCTT AAGAAGCGG CAGTTGCTG GAAGCCTGCA GAGCAAGGA AGCAGCGTTT 180
 GATGCCGAAT ATCAGCGAAA TCTTCACAGG GTGGACCTG ATATTTTAAC CTTTACGATA 240
 45 GCTCTGACT CTTCTAAGT TATCAACCT CTGATAGAAG AACTTGCTG CCATAAGTTT 300
 ATCAATAGAG AATAGTTAGT TGSTGACACT ACTTCAAGAG AACCTCTGCA TTCCAGTCAT 360
 ACCAATCTG CAACTTGATT TTCAGAAGTC AAGAGTATAT CGCGATAAGA CAGTGACAG 420
 50 GTGGAGGGGA AAAAAAGGG GAGGGGAAG CTTATCTGA AAAAGCATCA CAGAASTAGA 480
 AAAAAATGT GAAAGCATT TAACTSTAAC GTTCTTTGAG TTTGTGATT ATCCACATTT 540
 55 TTCCCCCTG ATTATGGAAA ATGTCTCTCA GCATTGCTTT ATTACAAAGT AAAGGATGGT 600
 TTTATAAAT TGAGACTGAT GAAACATCAA TACTAGAGCC CATGAGGATG AAAGAAATTA 660
 60 TCAAATAGTG CTGAACAGAA TAAGATGTTA ACGCTGAGTT ATTAGGACTG GAAGGCTATG 720

	GCAGTTGAAG GCGGAAGGCT CCACTGCATT CTTTGGCTAA GGCTGAATG CTGCTCATC	1140
5	TGTAAGATCT AIACTCGAGG TTTTGTMTTC CTTTAAAAAT TCTTTAGGGA GAGAGGGATG	1200
	GTTCCTGAGG GCTTCTGAAA GTATGATPCA ATCTGCAACA TACAGGTAGG TCTTCAGCAT	1260
	AAGCTGAAAT ATATGCATCT AAAAACTTTG ACATCTTTTT TTMTAATTTT CCACTTTCTT	1320
10	CTTAACTTTA CTCTCTTTT TGTCCCCCCC CCATCTTACA GAAGTTGAGG CCAAGGGAGA	1380
	ATGTTAGGCA CAGAAGAAAC ATGGCAAACCT GCTCTSTCT TTCAAACCAA AGTGTTCCTC	1440
15	CCAACCCCAA ATTTCTCTAA GCACTGSCCA GTCTTTTGT GGCATTSTTT TCTACAACCA	1500
	AATTCTGGGT TTTTCTCTTC TTCTTTTAAA CATAGAGTA CCACCACAAG GATGCCCTA	1560
	CTCTCTGGA GTCTCTGAAA GCATCTGTTT GAGGGAAGG TCTCTGGGA AGCAAGTGT	1620
20	TATTTGGAAT GCTTGTCTCC CTTTTCAC CTGGGACATT GYAATCATAA AATAACAGTA	1680
	AATTCACAAAC CTCAAAAACT ATTATGGCCT GAGCAGAGCT GAAATCTAGC AGAGTTTAAAC	1740
25	TCTTCTGCT CCATGCTCTGT CACTTATAAT TCAAGTTCTG CTGTTGGCTT CAGAAATGA	1800
	GCAGAAGAAAT CTTTTATGC TAGTTATGTC ATTCATGGTT GAAACTCAAC TTAGGGAAG	1860
	GTTTCCAATG TATTAAGCAA TGGGCTGCTT CTCCCCAATC CTCCCTAACA ATTCTGTGTG	1920
30	TGACATTCTC ATCTAAAAGG TTAGTGGCTT TTGCTTGGGA TCAGTGTCTCT CTATTGATGT	1980
	TCTTGTCTGT CTCCAGACAC ATTCTGTGTG CATTAAAGAT TGAAAGACTT GTAGATGTGT	2040
35	GATGTTCAAG CACACGATGC TGAAAGCTAT GTTACTATTC TTAGTTGTA AATTGTCCTT	2100
	TGATACCAT CATCTTGTTC TCTTTTGTG GTATAAATA AAACACTGT TGACAATAAA	2160
40	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	2219

(2) INFORMATION FOR SEQ ID NO: 104:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1351 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

	CTTCACAGAC TGACAGAATG GTTTTGTTCCT GTTTTGTTCCT GTTTTGTTCCT GTTTTGTGAGA	60
55	TGGACTCTAG CTCTGTGACC CAGGCTGGAG TGCAGTGGTG CGATCTCGGC TCACTGCAAG	120
	CTCCGCTCTC CGGTTTCTCA CCATTCTCCT GCTCAGCCT CCGGAGTAGT TGCGACTACA	180
60	GGCGCCACAC ACCAGGCCCG GCTAATTTTT TGTATTTTTT AGTAGAGAGG GGTTTTCACC	240

355

	ATGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCCTGATC CGCCCGCTC GGCTCCCAA	300
	AGTGCTGGGA TTACAGGCGT GAGCCACCGT GCTGCCCCA GAATGGTTTT TAAAGCCACA	360
5	GTTGAGARGC CACCCATTGC CGCCCGCTG GACAGTGATC ATCTTGTICA TCTTGTTCAG	420
	TCTTTCTTG TGTGATTGSA ATATTTCATC CCGTTTGAAA GATGAGAAGC TTGAGATGCA	480
10	AAGACTCTAC CTTTCCAAT TCTCACTGCT GSAAGARCT AGAAGCACAG TTCAAAGTTC	540
	TGNTTCTGG ACTCTGCAAT CCAGGTYTCC GTTGTCCAC TTGCTATCC TCAATGCCAC	600
	ACTGTTTTTG AAGTGGCCCA TAACTGSAAG GRAAAGTTTA AAGACAGTTC AATTAAATCA	660
15	TCAGRATSCA TTCTTTTTT TTTCGGAAC GGAATTTTAC TCTTCTGCG GAGCTGGAG	720
	TGCAATGCTG CAATGATCTC GCTCACTGC AACCTATGCC TCTTGGTTC AAGNGATTAT	780
20	CCAGCTCAG CTTCCGAGT AGCTGGGATT ATGGGCGGCC ACCACCTATC CCAGCTAATT	840
	TTTGATTTTT TTTTTFAGT AGAGATGGG TTTCGCCAGG TTGGCCAGG TGTCTTGTG	900
	AATCTCTGGC YTCAGTGAT YTGCCACYT CATCTCCAA AAGTCTGGG ATTACAGGCA	960
25	TGAGTCACTG CGCCTGGCT CAGAATGCAT TCTTACACAT CTATCCTAGA CATTTATAAG	1020
	CACTCTAATG GATAACAATC CAAGAAATAA TGATTGTAAA AGATGATGCC GAAGAGTTGA	1080
30	TGTCAATCTT TTTTTCCTAA GAAAAAAGT CGCGAGTAT TAAATATTTA GATCAATGTT	1140
	TATAAAATGA TTACTTTGTA TATCTCATT TTTCTATTTT GGAATAAAAA CTGACCTTCT	1200
	TTAATCATAT ACTTCTTTT TGTAATATAG AGCTTTTNTG TCATCTTCCC CACTTTATTA	1260
35	GTAAATTTAA ATTGGAAAAA ACTCTCAAAC TAATATTCTT GTCTGTTCCA GTCTTATAAA	1320
	TAAACTTAT AATGCATGTA AAAAATAAAA A	1351

40

(2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3066 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

	GGGACGAGGC GGTGAGGCG CAATATCACA GTTCCGGGCA TTGGGGGAAC CCGAGCCGCT	60
	TTGCTATGCT TATGCTATGCT TATGCTATGCT TATGCTATGCT TATGCTATGCT	120
	TTGCTATGCT TATGCTATGCT TATGCTATGCT TATGCTATGCT TATGCTATGCT	180
60	TTAAAAATAGC TCTTCTTATT ATAAAAATA TAAAAATAA GAATCTTTTA ATTCTTCTG	240

	GGATTCTGCT CGTGTTCACAA ATCATCGGCT TTCTGTGCG AGGCTTGATT GTTCCAGGCG	360
5	CCACAAGGCG AGTGTGCTAC ATGTGCTGA AATGTGTGA TGCCCGTAAG AAGCATCACA	420
	AGACAAAATG GTTGTGTGCT TGGGACCCCA ATCATGTGA CAAGATCCCA GACATTGAAG	480
	AGGCAATTCC AAGGGAATTT GAAGCCAAAT ACATGTGTT TTGTGTTCAC ATTCCGCTCC	540
10	CCACATGGA GATGAGTCTT TGGTTCACAT TCATGCTGTT TATGCTGAG CTGGACATTC	600
	CCCTTAAGCT AAACAACCAA ATCAGAGAAA ATGAGAAAT CTCATGAGAC GTTTCGCTCG	660
15	CTTACGCTGA TGAGGCATTT GTGAGTGA CTGAAATGCG CCAATGAAAG GTACCAAGGA	720
	AACTGAAATG CAGCTTCACA TCTCCCAAGA CTCCAGAGCA TGAGGGCGCT TACTATGAAT	780
	GTGATGTCTT TCTTTTCATG GAAATGCGT CTGTGCGCA TAAATTTTAC GTTTAAACA	840
20	TCCGGCTGCG TGTGAATGAG AAGAAAGAAA TCAATGTGG AATTGGGAG ATAAAGGATA	900
	TCCGCTGCT GGGGATCCAC CAAAATGAG GCTTCACAA GTGTGCTTT GCAATGAAJA	960
25	CTTGTCTTAC GTCCAGCATG TGCATCATT TGGTGTGTA TTGGAGGAGG ATCAACATCA	1020
	TGTCCCGACC CCGAGTCTTT CTGGAAAAAG TCATCTTTGC CTTTGGGATT TGCATGACCT	1080
	TTATCAATAT CCGAGTGAAG TGGTTTTCCA TGGGTTTGA CTGGACCTGG ATGCTGCTGT	1140
30	TGTGTGACAT CCGAGGCG ATCTCTATG CGATCTTCT GTCTTTCTGG ATCATCTTCT	1200
	GTGGCGAGCA CATEATGAT CAGCAGGAG GGAACACAT TGGAGCGTAT TGGAGGCAAG	1260
35	TGGAGCCAT TGGGTTTGG TCTTCTGCG TCTTCATATT TGACATCTGT GAGAGAGGGG	1320
	TACAAGTAC GAATCCCTTC TACAGTATCT GGAATACAGA CATTTGGAACA GAGCTGGCCA	1380
	TGGGCTTCAT CATGCTGCTT GGAATCTGCG TGTGCTCTTA CTCTCTGTTT TATGCTTCA	1440
40	TGGTATTTCA GTGTTTTCGG AACATCAGTG GGAAGCAGTC CAGGCTCCCA CCTATGAGCA	1500
	AAGTCCGCG GTTAACTAT GAGGGGCTAA TTTTATGTT CAAGTTCTTC ATGCTTATCA	1560
45	CTTGGGCTG CGGTGCCATG ACTGTGATCT TCTTCATCT TAGTCAGGTA ACGGAAGGCG	1620
	ATTGGAATG GGGGCGGCTC AAGTCCAAAG TGAACAGTGG CTTTTTCAAA GGCATCTATG	1680
	GEATGTGAA TCTGTATGTC TTGTCTCTGA TGTCTTTSTA TGCACCATCC CATAAAAACT	1740
50	ATGGAGAAGA CCACTCCAAT GGAATGCAAC TCCATGTAA ATCGAGGGAA GATTGTGCTT	1800
	TGTTTGTTC GGAACCTTAT CAAGAATTGT TCAGGCTTC GAAATATTCC TTCATCAATG	1860
55	ACAACGCAGC TTCTGGTATT TGAGTCAACA AGGCAACACA TGTTCATCAG CTTTGCATTT	1920
	GCAGTTGTCA CAGTCACATT GATTGTACTT GTATACGCAC ACAAATACAC TCATTTAGCC	1980
	TTTATCTCAA AATGTTAAAT ATAAGGAAAA AAGGCTCAAC AATAAATATT TTGAGTATA	2040
60	AAAAAAAAA AAAAAAAAAA AAAAAA	2066

5 (2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 1705 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15 AATTGGCAK AGGCCAGCTG TGGCTGGAA GGAAGTGGTC TGCTCACAAT TGCTGGCTTG 60
CGCATCAGGA CTGCTTTTAT CTCTGACTC ACGGTGCAAA GGTGCACTCT CGGAACGTTA 120
AGTGGTGGC CAGCGCTTGG AATCTACCG CCCCCACAGC CGGATCCCT CAGCCCTCCA 180
20 GGTCTCAAC TCGGAGGAC GTGAACAAT GGCTCCATG GGGCTACAG TAATGGGCAAT 240
GGGCTGGCG GTCTGGCT GTCTGGCT CATCTGTG TGCGGCTTC CATCTGGCG 300
25 CTGACGGCG TGTATGGCA GGAACATCT CAGCTCCAG ACCATCTG AGGGCTTATG 360
GATGAAGTGT GTCTGGCA GGAACGGCA GATCAAGTGT AAGGTGAGG AATGCTGCT 420
GGCACTGGC CAGGAGCTGC AGGGGGCGG GGGCTGTG ATCATAGCA TATCTGGC 480
30 TGCTGTGGG GTCTCTCT CTCTGTGG GGGCAAGTGT ACCAAGTGT TGGAGGATGA 540
AAGCGCAAG GGAAGACA TATCTGTG GGGCTGTG TTCTGTGT CGGGCTTAT 600
35 GGTGATASTG GTCTCTCT GAGCGGCA CAGATCAT CAAGACTCT ACAATCTGCT 660
GGTGGCTTC GGGAGAGC GGGAGATGG TGTCTGCT TACGTGGGT GGGCTGCTC 720
CGGCTGTG GTCTGTGG GGGCTGTG TTGTGCAAG TGTCCACC TCACAGACAA 780
40 GCCTTACTCG GCAAGTATT CTGTGCGG CTCTGTCT GCGAGCACT ACGTGAAG 840
TGGACGGCT CCACTCTCT CCTCTCTCT TTCTGTCT CTGAGTGA CTCAAGCAG 900
45 GGTGTGAGC CAGGAGGCG CTCTAGGAG CCACTGCT CTGGGAGT GGGCTGTG 960
AGAGACTGAG CAGGAGGAG AGGAGGAG CTCTAGGCT TGTGGGAG TGGAGCACT 1020
TCCCAAGGCG GCTCTCTGCT AGCAAGAACA GATCCACT TCTCTGAT ATTGGGAGG 1080
50 GACCGAAGT ACAGGCTGT GTCTGAGT GGGAGCTG CTCTGTCT CAGGATGC 1140
TTAAGCTGA CTCTGAGT TCTCTGAG GGTGTGCT ACTGTGCGA TTACATTT 1200

1260
60

CAGGTTTGGG CAGTGGTGGG GGAGGGGGCC AGAGAGGGGG CTCAGGTTCG CCAGTCTCTT 1440
 GGCTTCAGGA CTCTCTGCTT CACCGGTTTC AGCTCAGGGG CCGTGGAGAC TGATCCCTTC 1500
 5 TGAGTCCTCT GCGCCTTCCA AGGACACTAA TGAGCCTGGG AGGGTGGCAG GGAGGAGGGG 1550
 ACAGCTTCAC CCTTGAAGT CCTGGGTTTT TTCTCTTTC TTCTTTGTGG TTTCTGTTTT 1620
 10 GTAATTTAAG AAGAGCTATT CATCACTGTA ATTATTATTA TTTTCTACAA TAAATGGGAC 1680
 CTGTGCACAG GRAAAAAAAAA AAAAG 1735

15

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1167 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TGCAGGAATT CGGCAGAGGT TTTCGGTAG ATTCTGGCAG TTGGTGAGCA TCATGGCAAC 60
 CATTACAGCC ACAATCAAAG TCCCGAGAT CCGTATGTA ACAAGSATG AGCGAATGGG 120
 30 TCCCACTCC CACATCCGGG GACTGGGGCT GGACGATGC TTGGAGCCTC GGCAGGCTTC 180
 GCAAGGCAATG GTGGTCAAG TGGGGCAAG GGGGGGGCT GCGTGGTCC TGGAGATGAT 240
 35 CCGGSAAGGG AAGATTGGCG GTGGGCAAT CTTATTGCT GCGCAGGCG GCAAGGGGAA 300
 GACGCGCATC GCAATGGCA TGGGGCAAG CTTGGGCTT GACAGGCAAT TCACAGCCAT 360
 CCGCGGCAAT GAAATCTTCT CCGTGGAGAT GAGCAAGACC GAGGCGCTGA CGCAGGCTTT 420
 40 CCGGCGGTCC ATCGGCTTC GCATCAAGGA GGAGACGGAG ATCATCGAAG GGGAGGTGGT 480
 GGAGATCCAG ATTGATGAC CAGCAACAGG GACGGGCTCC AAGGTGGGA AACTGACCT 540
 45 CAAGACACA GAGATGGAGA CCATCTACGA CTTGGGCACC AAGATGATG AKTCTCTGAC 600
 CAAGGACAAG GTCCAGGCG GGACGTGAT CACCATGAC AAGGGGACCG GCAAGATCTC 660
 CAAGCTGGGC CGCTCTTCA CACGGGCCC CGAAGTACA CGCTATGGGC TCACAGACCA 720
 50 AGTTCTGCA GTGCCAGAT GGGGAGCTCC AGAAACGCA GAGGTGGTG CACATCGTGT 780
 CCTTGCACGA GATGAGCTC ATCAACTCTC GCACCCAGGG CTCTCTGGCG CTCTTCTGAG 840
 55 GTGACACAG GGAGATCAAG TCAGAACTCC GTGAGCAGAT CAATGCCAAG GTGGCTGAGT 900
 GCGGCGAGGA GGGCAAGGG GAGATCATCC CTGGAGTGCT GTTCATGAC GAGGTCCACA 960
 TGCTGGACAT CGAGAGCTTC TCCTTCTCA ACCGGGCCCT GGAGAGTGAC ATGGGCGCTG 1020
 60 TCCAGCAGGT CTATGGGAT GCCGTGAGGG CTCTGCTAGC TGGTGGCGCG GATTGGGCTG 1080

ATGCCACGGT TGGTGGGCTG GTGCGAATT CCTGCACCC GGGGGATCCA CTAGTTTATG 1140
 AGCGGGCCGC ACCGCGGTGG ANCTGDN 1167

5

(2) INFORMATION FOR SEQ ID NO: 108:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

20

GGCAACAGGG AATCATCGTG TGATGTGTGT CCTGCTTTT TGAGTGTGTG GAGTCCTGCT 60

CAGGTGCTAG GTACAGTGIG TTTCATCGTG GTGGCTTIGAG GGGAAACCTT GTTCAGAGCT 120

GTBACTGCGG CTGCACTCAG AGAAGCTGCG CTTCGCTGCT CTTAGGCGCG GGCCTTCTCT 180

25

CCTCCTCATG ATCCAGAGCA GTCAGTGTTC GGGAGGCAGA AGGTACCGGG GCAGCTACTG 240

GAGGACTGTT CGGGCCTGCG TGGGTGCGCG CCTCCGCGCT GGGGCGCTGT TGCTGCTGTC 300

30

CATCTATTTG TACTACTGCG TCCCAAATGC GGTCCGCGCG CCTTCACIT GGATGCTTGC 360

CCTCCTGCGG CTCTCGAGG CACTBAACAT CCTCCTGCGG CTCGAAGGCG TGGCCCCAGC 420

TGAGATCTCT GCAGTGTGTG AAAAAGGGA TTTCAACGTC GGCATGCGC TGGCATGGTC 480

35

ATATTACATC GGATATTTTC GCGTGATCCT GCCAGAGCTC CAGGCCCCGA TTCSAACTTA 540

CAATCAGCAT TACAACAATC TGCTACCGCG TGCAGTGAGC CAGCGGCTST ATATTCTCCT 600

40

CCCATTGAGC TGTGGGCTTC CTGATAACCT GAGTATGCTT GAGCCCAACA TTGCGTTCTT 660

GGATAAAGTG CCCCAGCAGA CCGGTGACCG TGCTGGCATC AAGGATCGGG TTTACAGCAA 720

CAGCATCTAT GACCTTCTTG AGAACGGGCA GCGGCGCGCG ACCTCTCTCT TGGAGTACGC 780

45

CACCCCTTTC CAGAGTTTCT TTGCTATGTC ACAATAAAT CAAGTGGCT TTAGCGGGGA 840

CGATAGCTTT GAGCAGGCA AACCTTTCTG CCGGACACTT CAGGACATTC TGCCASATGC 900

50

CCTTGASTCT CAGAACAAT CCGGCTCAT TGCTACGAG GAACCTGCAG ATGACAGCAG 960

CTTCTCTGTC TCCCAGGAGG TTCTCCGGCA CCTCCGCGAG GAGGAAAAGG AAGAGGTTAC 1020

TGTGGGCGCT TTGAAGACT CAGCGGTGCG CAGTACCTCC AGGATGTGCG AAGAGGCTGA 1080

60

CTGCTTATCT CAGGAGCA AATATTCTCT TGCATGAG AGTTTCTGCT TAAAGCTTCA 1140

360

GGACTTGACA TCTTAAGATG CGTCTTCTTC CCTTGGGACA GTCATTTTCC CTCTCTGAGC 1320
 CTGGGTCTCT TCAACCTGTG AAATGGGATC ATAATCACTG CCTTACCTCG CTCADGGCTG 1360
 5 TTCTGAGGAC TGAGTGTGTG GAAGTTTTTC ATAAACTTTG GATGCTAGTG TACTTAGGGG 1440
 GTCTGGGAGG TCTCTTCTAT GGGGGCTTCC AGACCCATTC CCGACCTTCC TGGCTTCTCT 1500
 10 TTGGGGGGGG ACGCCGAACCT CTCTCAATGG TATCAACAGG CTGCTTGGCC CTCTTGGCTCC 1560
 TGCTCATCTT CCATTATTTG GGAGGGGAGC CAGAAGAATG GAGAGGAGGA GGAGGCTGAG 1620
 TTTGGGTAT TGAATCCCCC GGTTCGCCAC CTGAGGATC AAGTTTGTA TGGACTCTCC 1680
 15 TGCGGGGCAA CTCTTGOSTA ATCATGACTA TCTTAGGAT TCTGGCAGCA CTCTCTTCCC 1740
 TGGCTCTTA AGCTAGCTG TGTATGGGCA GGGGACCCC ACTAGAGTAC TGGCTCTCAC 1800
 TTGGGCTTTC CTATACTCT ACGCTTTCT CAACTGCTCT TTTTAAAGT ACATCTCAGA 1860
 20 TTAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGG GGGGGG 1900

25

(2) INFORMATION FOR SEQ ID NO: 109:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 611 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

40

ATGAATTAAC GCGAAGCTNT NAATAGGGAC TCACTATGGG GGAAAGNTGG GTAAAGCCTG 60
 CAGGTACCTT TCGGAATTC CGGGGTGAGC CCAAGCGTCC GATGGGGCTT TACTAAATCA 120
 40 GGCTTGCAAG CTCAAAGCTG CAATCTGCCC ACTCTCAGST ACTGAGACTT TGTGGGCTC 180
 AGACACCAGG AAGAAAGTTG GGATACAGTC ATTTGAGTTA AAAAGGGAAT GACCCCTCAG 240
 45 AAACCCGCAT TACCACTCTT ACTCTTGAA GTGCTTTAC TTTTAACGCT CTCTGTTCTG 300
 AAAAAGAGST GTTTGGTTAC GTGTAGGCA ACATCAAGTT TTGTTAGCTG TGATTACCT 360
 TTGTCCGTTT AAAAGACTTC ACGGAGCCAT TCTGTATACA AGGTGTGCTC TTCCCAATGT 420
 50 AGAAGGGGTT ATGGAAAAGG GTGGATCCT TTGCTGTAAA CTGGAGAGAC CAGTCCCAA 480
 CAGAGGGGAA TTTTAAGCCC TTCTCATCAC CCAATTGGAT GTTTTGTCTT ATAGCAAATT 540
 55 CCTGCAAAAT AAATAAATAA ATATTTGCAA AACTAAAAA AAAAAAAAAA AAAAAAAAAA 600
 GGGGGGNCN C 611

60

(C) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2632 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

10 TCCGAGCTCT CAGGACAAGG GGCCTGGGCG ATCTTTTAAA AAAGCCGATT GGGTGTCTTT 50
 CTAAAANTAC AACCAGTACT TCATCGTCAA GTTCTCGGGA AGGAGTCCC CTCCAGATTC 120
 15 TCATGGASTG ACAAATCTTG ACTCTTCTTC CTGGAATTTT TCAGGCCCAA ACTAGCGTTT 130
 CTACAATGAT TTAATTTGCA AATTTGTCTT GATTATGGGT GCGTGTGAG SAACGTGCTT 240
 TTGTTAGGAA CCGAACTGG GCGCGCTGA GGGCGTGTAC GCAATGASTC CGGAAGAGGG 300
 20 TGAAATGCTT TCGTAGGCA CTCCAGGCT GTGAAGATGG CGCGGCTGC GTGGTTTAC 360
 GTGTTGCTG TCAATCTCT GCTTCTGGA CCTCACCGT CACCACTGTG GTTTTTCAGT 420
 25 GCGGGACCGG CAACCGTAGC TGCTGCCGAC CGCTCCAAAT GGCACATTC GATACCGTCG 480
 GGGAAAAATT ATTTTAGTTT TGGAAAGATC CTCCTCAGAA ATACCACTAT GTTCTGAAG 540
 TTGATGAGG AACCTTGTGA CTTGTCTTTG AATATAACCT GGTATCTGAA AAGCCTGAT 600
 30 TGTTACAATG AAATCTATAA CTTCAAGGCA GAAGAAGTAG AGTTGTATTT GAAAAACTT 660
 AAGGAAAAAA GAGGCTTCTC TGGAAATAT CAAACATCAT CAAAATTGTT GAGAACTGC 720
 35 AGTGAACCTT TAAAACACA GACCTTTTCT GGAGATTTTA TGCATCGACT GCTCTTTTTA 780
 GGAGAAAAAC AGGAGGCTAA GGAGAATGGA ACAAACCTTA CCTTTATTGG AGACAAAACC 840
 GCAATGCATG AATCATTGCA AACTTGGCAA GATGCACCAT ACATTTTTAT TGTACATATT 900
 40 GGCATTTTAT CCTCAAAGCA ATCATCAAAA GAAATTCAC TGAGTAATCT TTTTACCATG 960
 ACTGTTGAAG TGAAGGGTCC CTATGAATAC CTACACTTG AAGACTATCC CTTCATGATT 1020
 45 TTTTTCATG TATGTGTAT TGTATATCTC CTGTTTGGTG TTCTGTGGGT GGCATGCTCT 1080
 GCTGCTACT GGAGAGATCT CTGAGAGATT CAGTTTGGGA TTGGTGCTCT CATCTTCTG 1140
 GGAATGCTTG AGAAAGCTGT CTCTATGCG GAATTCAGA ATATCCGATA CAAAGGARAA 1200
 50 TCTGTCCAGG GTGCTTTGAT CTTTGCAGAR CTCCTTCAG CAGTGAACG CTCACTGGCT 1260
 CGAACCTTGG TCATCATAGT CAGTCTGGGA TAAGGCATG TCAAGCCAG CCTGGACTCA 1320
 60 TGTAGAGAT GCTTTTCTT TGTGTGATAT TTTTATAT GATTAAGCA ATTAAGCTAT 1380

362

TAAAACTTGG GAGGAACATT GTAAAACTCT CTTTGTATGG GCATTTTCACC AACAGGCTTA 1560
 TTTTGGCACT GGCAGCATCC ATTGTGTTTA TCATCTGGAC AACCATGAAG TTCAGAATAG 1620
 5 TGACATGTGA GTGGACTGG CGGGAGCTGT GGTAGACGA TGCCATCTGG CGCTTGCTGT 1680
 TCTCCATGAT CTTCTTTGTC ATCATGGTTC TCTGGGAGC ATCTGCAAAE AACCAGAGGT 1740
 10 TTGCTTTTC ACCATTGTCT GAGGAAGAGG AGGAGGATGA ACAAAGGAG CCTATGCTGA 1800
 AAGAAACCTT TGAAGGAATG AAAATGAGAA GTACCAACA AGAACCCAAT GGAAATAGTA 1860
 AAGTTAACA AGCAGAGGAA GATGATTTGA AGTGGGTAGA AGAGAATGTT CCTTCTTCTG 1920
 15 TGACAGATGT AGCAATTCCA GCCCTGTGAG ATTCAAGTGA GGAACGAATG ATCACACACT 1980
 TTGAAAGGTC CAAAATGAG TAAGGAATGG GAAGATTTCG AGTTAAAGAT GGCTACCATT 2040
 20 AGGGAAGAGA TCAGCATCTG TGTAGTCTT CTGTAAGGCT CCATGGGATT AAAGGAAGCA 2100
 ATGACATCTT GATGTGTTCC TTGATTTTTG GGCATTGGAG TTGGGGAGAG GTGTCAGAAC 2160
 AAAGAGACA TCTTACTGAA AACAAGTTCA TAAGATGAGA AAAATCTACG AGCTTCTTAT 2220
 25 TTACAACACT GTTGGCCCTT TTCCTCCAG ACTCTGACAT GGATGTTTAT GCAACTTAAG 2280
 TGTGTTGTTG CTGAACTTTC TGTAAATGTTT CATTTTTTAA ATCTGACAAA CTAAAAAGTT 2340
 30 TAAGTCTTTC TAAAGATTG TCATCAACAC CATAATATGT AATCTCCAGG AGCAACTGCC 2400
 TGTAATTTTT ATTTATTTAG GGAGTTACAT AGGTGATGGG GGAAATTGTT AACTACCTTT 2460
 CATTTTCTG GGAAGTCAAG GTTACATCTT GCAGAGGTG TTTTGAGAAA AAAGGGCCCT 2520
 35 TCTGAGTTAA GGAGCCATAG TTCTATCAAT GATCAAAAGA AAAAAAAAAA AACTCGATCG 2580
 GCACGAGGGG GGGCCCGSTA CCCAATTCGC CCTATGGGAN TCGAATGAGA CC 2632

40

(2) INFORMATION FOR SEQ ID NO: 111:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2249 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GAATTGCGCA CGAGCTCACC GTGCTGCGTG ACACAAGGCC AGCCTGCGCC TACGAGCCCA 60
 TGGACTTTKT RATGCCCTC ATCTACGACA TGGTACTGSW TGTGGTCACC CTGGGGCTGG 120
 55 CCTCTTCAC TCTGTGCGGC AAGTTCAAGA GGTGGAAGCT GAACGGGGCC TTCTCCTCA 180
 TCACAGCCTT CCTCTCTGTG CTCATCTGGG TGGCCTGGAT GACCATGTAC CTCTTCGGCA 240
 60 ATGTCAAGCT GCAGCAGGGG GATGCCTGGA ACGACCCAC CTTGGCCATC ACGCTGGCGG 300

	CCAGCGCTGG GTCTTCGTCA TCTTCCACGC CATCCCTGAG ATCCACTGCA CCGTTCTGCC	360
5	AGCCCTGCAG GAGAACACGC CCAACTACTT CGACACGTGG CAGCCCTAGGA TGCGGGAGAC	420
	GGCCTTCGAG GAGGACGTGC AGCTGCCGCG GGCCTATATG GAGAACAAAG CCGTTCTCCAT	480
	GGATGAACAC AATGCAGCTC TCCGAACAGC AGGATTTCCC AACGGCAGCT TGGGAAAAAG	540
10	AACCAGTGGC AGCTTGGGGA AAAGACCCAG CGCTCCGTTT AGAAGCAAAG TGTATCAGCC	600
	AACTGAGATG GCCGTGGTGC TCAACGGTGG GACCATCCCA ACTGCTCCGC CAAGTCACAC	660
15	AGGAAGAMAC CTTTGGTGAA AGATTTTAAG TTCCACAGAA TCAGAATTTG TCTTACCGAT	720
	TTGCCTCCCT GGCTGTGTCT TTCTTGAGGG AGAAATCGGT AACAGTTGCC GAACCCAGGC	780
	GCCTCACAGC CAGGAAATTT GGAATCCTA GCCAACGGGA TTGTGTGTA ATGTGAACAT	840
20	TGACGAACTG AAAAGCTAAC ACCGACTGCC CGCCCTCCC CTGCCACACA CACAGACAGC	900
	TAATACCAGA CCAACCTCAA TCCCCGCAA CTAAAGCAA GCTAATTGCA AATAGTATTA	960
25	GGCTCACTGG AAAATGTGTC TGGGAAGACT GTTTCATCCT CTGGGGGTAG AACAGAACCA	1020
	AATTCACAGC TGGTGGGCA GACTGGTGTG GTTGGAGGT GGGGGGCTCC CACTCTTATC	1080
	ACCTCTCCC AGCAAGTGT GCACCCACAG TACCTCTTG GAGATGACCG TTGCTTTGAG	1140
30	GACAAATGGG GACTTTGCCA CCGGCTTTGC CTGGTGGTTT GCACATTTCA GGGGGGTGAG	1200
	GAGAGTTAAG GAGGTTGTGG GTGGGATTCC AAGGTGAGGC CCAACTGAAT CGTGGGTGGA	1260
35	CGTTTATAGC CAGTAGAGT CGAGGGACCC TGGCATGTGC CAAAGAAGAG GCGCTCTGGC	1320
	TGATGAAGTG ACCATCACAT TTGAAAGTG ATCAACCACT GTTCTTTCTA TGGGGCTCTT	1380
	CGTCTAGTGT CTATGCTGAG AACACAGGCC CGCCCCCTTC CCTTGTAGAG CAGTAGAAAT	1440
40	ATTCTGGCTT GGGGCAGCAG TCCTTCTTTC CATTGATCAT CTCGCCCTGT TCTACACTT	1500
	ACGGGTGTAT CTCCAAATCC TCTCCCAATT TTATTCCTTT ATTCAATTTCA AGAGTCCAA	1560
45	TGGGGTCTCC AGCTGAAAGC CCGTCCGGA GGCAGTTGG AAGGCAGGCA CCAACCCAGG	1620
	TTTTCCGCGA TGATGTCACC TAGCAGGCT TCAGGCTTC CCACTAGGAT GCAGAGATGA	1680
	CCTCTGGCTG CCTCACAAGC AGTGACAGCT CCGGTCTTTT CCGTTGCTAT GGTGAAAATT	1740
50	CCTGGATGCA ATGATCACA TGAGGTTTC TTGTTCTTT TGGAGGTTCT GGGGATATT	1800
	TTGTTTGGT TTTCTGCG GTTCATGAA AACAGCCCTT TTCCAGGCC ATTGTTCTG	1860
	TTTAAATG ACTTTAA TCACTGCTGCTTAA ATAAAGTCTT TTTT	1920
60	AGATACTCTA ATCACTACAT TGTCTTTCT ATAAACTAC CAGTAAGCTT TTATCTTTA	2100

AAGAAAAATG AAAAAGCTTA GTTTTGGGGG GCGGGGGGAG GACTGACCCG TTCATAAGCC 2160
 AGTACGTCTG AGCTGASTAT GTTCAADAA ACCTTTTGAT ATTTCTCAAA AAAAAAAAAA 2220
 AAAAAACCTG GGGGGGGGGG GGAAGCTGG 2249

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2198 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GATACTATAA GGGAAATGAC TCAGGGGTGC GCGGTAGAC TAGTGGATCC CGGTTGCAGG 60
 AATTGGCCAG AGGGGGGGGG GAGCCGAAGT GGTGGGGGCC CCGGGGCGCG TGGTTCGGG 120
 GANCCCCAAA TCTGAAATG CAGCTTGAAG ACCCCGAAGA AAAGSAGSAA TTGGCGGTGC 180
 CCGAGAATAG CTGGCTCCAG CAGTTTAAGG AAGAAATGTC TAAAGTTTTT AAATCADATA 240
 CTGACCAACT TGTGTTGATA TTGCTGCAA AATTTTGA AATCAAGAT AATTGAGTC 300
 AGCATGGAAT TCATGATGGA CTACTGTTG ACCTTGTGAT TAAACACAA AACAGGCTC 360
 AAGATCAGTC AGTCAGCA ACAAATACAG CTGGGAAGCA TGTTACTACA TCACTAACTC 420
 CTAATAGTAA CTCTACACTT GCTCTGCTA CTAGCAACCC TTTTGGTTTA GGTGGGCTTG 480
 GGGGACTTGC AGTCTGAGT AGCTGGGTT TGAATACTAC CAACTTCTCT GAACTACAGA 540
 GTCAGATGCA GGCACAACTT TTGCTTAACC CTGAAATGAT GGTCCAGATC ATGAAAAAAG 600
 CCYTTGTGCA GAGCATGCTC CTCAAATGCT GACCTGATGN AGACAGTTAA TTATGGCCAA 660
 TCCACAAATG CAGCAGTTGA TACAGAGAA TCCAGAAAT TAGTCATATG TTGAATAATC 720
 CAGATATAAT GAGACAAAG TTGGAAGTTG CTCAGGAATC CAGCAATGAT GAGGAGATG 780
 ATGAGGAAGC AGGACCGAGC TTGAGCAAC CTAGAAAGCA TCCAGGGGG ATATAATGCT 840
 TTAAGGGCCA TGTACACAGA TATTCAGGAA CCAATGTGGA GTGCTGCACA AGAGCAGTTT 900
 GGTGGTAATG CATTTGCTTC CTGCTGAGC AATACATGCT CTGCTGAAGG TATCAACCT 960
 TCCCTACAG AAAATAGAGA TCCACTAGCC AATCCATGGG CTCCACAGAC TTCCAGAGT 1020
 TCATCAGCTT CAGCGGCAC TGCAGCACT GTGGGTGGCA CTAAGGTAG TACTGCCAGT 1080
 GCACTTCTG GGCAGAGTAC TACTGCCCA AATTTGTGCT CTGAGTAGG AGCTAGTATG 1140
 TTCAACACAG CAGGAATGCA GAGTTGTTG CAACAAATAA CTGAAAAACC ACAACTTATG 1200

365

CAAAACATGT TGTCTGCCCC CTACATGAGA AGCATGATGC AGTCACTAAG CCAGAATCCT 1260
 GACCTTGCTG CACAGATGAT GCTGAATAAT CCCCTATTTC CTGGAAATCC TCAGCTTCAA 1320
 5 GAACAAATGA GACAACAGCT CCCAACTTTG CTCCAACAAA TGCAGAATCC TGATACACTA 1380
 TCAGCAATBT CAAACCCCTAG AGCAATGCAG GCCTTGTTAC AGATTGAGCA GGGTTTACAG 1440
 ACATTAGCAA CGGAAGGCTT GGGCTCATC CCAGGGTTTA CTCTGGCTT GGGGGCATTA 1500
 10 GGAAGCACTG GAGGCTCTTC GGGAACTAAT GATCTAAGC CCACACCTAG TGAACACACA 1560
 AGTCCACAG CAGGAACAC TSAACCTGA CATCAGCAST TTATTCAGCA GATGCTGCAG 1620
 15 GCTTTGCTG GASTAAATCC TCAGCTACAG AATCCAGAAG TCAGATTTC GCAACAACTG 1680
 GAACAACTCA GTGCAATGGB ATTTTGAAC CGTGAAGCAA ACTTGCAAGC TCTAATAGCA 1740
 ACAGGAGGTG ATATCAATCC AGCTATTGAA AGGTTACTGG GTTCCAGCC ATCATAGCAG 1800
 20 CATTTCTGTA TCTGAAAAA ATGTAATTTA TTTTGTATAA CGGCTCTTAA ACTTTAAAT 1860
 ACCTGCTTTA TTTCAATTTG ACTCTTGGA TTCTGTGCTG TTATAAACAA ACCCAATATG 1920
 25 ATGCATTTTA AGGTGGAGTA CAGTAAGATG TGTGGGTTTT TCTGTATTTT TCTTTTCTGG 1980
 AACAGTGGGA ATTAAGCCTA TGCATGCAT CACTTCTGCA TTTATTGTAA TTTTITAAAA 2040
 ACATCACCTT TTATAGTTGG GTGACCAGAT TTTGTCTTCC ATCTGTCCAG TTTATTGCT 2100
 30 TTTTAAACAT TAGCCTATCG TAGTAATTTA TGTAGAATAA AAGCATTAAG AAGAAGCAAA 2160
 AAAAAAAAAA AAAAAATCTT GCGCCCGCGA ATTCTTCT 2198
 35

(2) INFORMATION FOR SEQ ID NO: 113:

40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1043 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

CTGAAGTSTA TGTGCTGAGG AAGAAGAGGC TCCTACTSTA GACAGCCTTG TTCTACAGAT 60
 50 CCTCCAGAAA ATCTCTGGAC CAGGTGGAAC CCAGGCTCAG AGAGGATGG GAGAGAGTT 120
 TAATTTTCCA TGATAAATAA AATCTATAA AATAATAAAC AAGAGAAAAG AGATTGAAA 180
 60 TGGCTTAAAT CAGGCGAGTC TCATCAGTGG CTGTGACTTG GCGGAGGTC TTAGGCTTGA 420

5 RGACTTGGAT GGGTTTGAGG GTTACTCOCT GAGTGAAGTG CTGTGCCTGG CTTTGTGGA 480
 AAGCAAGTTC AACATATCAA AGATWAATGA AAATGCAGAT GGAAGCTTTG ACTATGGSET 540
 CTTCAGATC AACAGCCACT ACTGCTGCAA CRATTATAAG AGTTACTCGG AAAACCTTTG 600
 CCACGTAGAC TGTCAAGATC TGCTGAATCC CAACCTTCTT GCAGGCATCC ACTGCGCAA 660
 10 AAGGATTGTS TCCGGAGCAC GGGGATGAA CAACTGGSTT AGAATGGAAK KTTGCACTGT 720
 TCAGGCCGGC CACTCTTCTA CTGGCTGACA GGATGCCGCC TGAGATKAAA CARGGTGCGG 780
 GTGCAAGGTS GATTCATTCG AAGACTCTTG TCTCACTCA RGGATTCTTC ATTTCTTCTT 840
 15 CCTACTGCTT CCACTTCATG TTAATTTCTT CCGTCCCAT TTACAACTAA AACTGACCAG 900
 AGCCCCAGGA ATAAATGGTT TTCTTGCTT CTCTCTACT CCGATCTGTA CCGAGTCCGC 960
 20 TGGTTCTGT CTGTTATTTG TAAACTGAGG ACCACAATAA AGAAATCTTT ATATTTATCG 1020
 AAAAAAAAAA AAAAAAACT CGA 1043

25

(2) INFORMATION FOR SEQ ID NO: 114:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 703 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

GAATTGCGCA CGAGTGGCGG GGCACCACGG CGSTTTTTTG ACGTGGCGG TGGACGCAGG 60
 CAGCATGGAC CACGGTTGCT GGGCGGATGG GGAGCGTCTA TGGTCAGTTG CTTAGAAGT 120
 40 GGTGAGATGG GAAGCTGCAG TTGGAAGACC CTGGAGGATG CCTGACAAAG GSATGTCTGA 180
 CACATGATTG GAGCTCTTTT TGAAATGTTT CTTGCCCTTC CTGGAGCAGA GGAGCCATTA 240
 45 TTTATGCAGG TACATCGAAG TCTTTTGACC TCCATACAGT GATTATGCTT GTCATCGCTG 300
 GTGGTATCCT GCGGGCCTTG CTCTGCTGA TAGTGTGCTT GCTGTGTCTT TACTTCAAAA 360
 TACACAACGC GCTAAAAGCT GCAAAGGAAC CTGAAGCTGT GGCTGTAAAA AATCACAACC 420
 50 CAGACAAGST GTGGTGGGCG AAGAACAGCC AGGCCAAAAC CATTGCCAAG GAGTCTTGTC 480
 CTGCCCTGCA GTGCTGTGAA GGATATAGAA TGTGTGCCAG TTTTGATTC CTGCCACCTT 540
 55 GCTGTTCGCA CATAAATGAG GGCCTCTGAG TTAGGAAAGG TGGGCACAAA AATCTTCATG 600
 AGCAATACTT CTTAGTAGAT TGTTTTGTTA TTCAAATCAA GTTCTAGTGT TTTTATGTGA 660
 GATTATATAA TTTACAGTGT TGTTTTATAT ACTTTTGAAT AAA 703
 60

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

5 GGCAGAGGGG GCATGAGCAG GAGGAGGATT ACCGCTACCA GGTGCTCAGG CCGGAGCAGA 50
 TTCTACAACA CATGGTGGNA ATGTATCCGG GAGGTCAACG AGGTGATCCA GAATCCAGCA 120
 ACTATCACAA GAATACTCCT TAGCCACTTC AATTGGGATA AAGAGAAGCT AATGGAAAGG 180
 10 TACTTTGATG GAAACCTGGA GAAGCTCTTT GCTGAGTGTG ATGTAATTAA TCCAAGTAAA 240
 AAGTCTCGAA CACGCCAGAT GAATACAAGG TCATCAGCAC AGGATATGCC TTSTCAGATC 300
 TGCTACTTGA ACTACCTTAA CTCGTATTTG ACTGGCCTTG AATGTGGACA TAAGTTTGTG 360
 25 ATGCAGTCTT GGAGTGAATA TTAACTACC AAAATAATGG AAGAAGGCAT GGSTCAGACT 420
 AATTGCTGTC CTGCTCATGG TTSTGATATC TTAGTGGATG ACAACACACT TATGCCCTG 480
 30 ATACAGATT CAAAAGTTAA ATTAAAGTAT CAGCATTTAA TAACAAATAG CTTTGTAGAG 540
 TGCAATCGAC TGTAAAGTG GTSTCTGCCC CCAGATTGCC ACCATSTTST TAAAGTCCAA 600
 35 TATCCTGATG CTAAACCTGT TCGCTGCAAA TGTGGCGGCC AATTTTGCTT TAACTGTGGA 660
 GAAAATTGGC ATGATCTCTT TAAATSTAAG TGGTTAAAGA AATGGATTAA AAAGTGTGAT 720
 GATGACAGTG AAACCTCCAA TTSGATTSCA GCCAACACAA AGGAATGTGC CAAATGCCAT 780
 40 GTACACAATT AGAAGGATGG TGSTTGTAAT CACATGGTCT GTCGTAACCA GAATTGTAAA 840
 GCAGASTTTT GCTGGGTGTG TCTTGCCCA TGGGAACCA ATGGATCTGC CTGGTACAA 900
 45 TGTAACCGCT ATAATGAGGA TCATGCAAA GCAGCAAGA ATGCACAGGA CGGATCTAGG 960
 GCAGCCCTGC AGAGTACCT GTTCTACTGT AATCGCTATA TGAACACAT GCAGAGCCTG 1020
 CGCTTTGAGC ACAAACCTATA TCTCAGGTG AAACAGAAAA TGGAGGAGAT GCAGCAGCAC 1080
 50 AACATGTCTT GGATTGAGGT GCAGTCTCTG AAGAAGGCAG TTGATGTCTT CTGCLAGTGT 1140
 CGTGCCACAC TCATSTACAC TTATGTCTTC GCTTTCTACC TCAAAAAGAA TAACCACTCC 1200
 TAAGATCTCT GGGAGTACAT TGAGGACTGA GAATGCCCTT GCATAAAATG AACTGTGAAA 1260
 60 AAAGATCTCT GGGAGTACAT TGAGGACTGA GAATGCCCTT GCATAAAATG AACTGTGAAA 1320

	ACTTTACCAT CTAGAGTCT CATECAATTA AAACAAAACA AACACAAACA AGGAGGCACT	1500
5	AAGCCTATTC TGACACCACT GGTGTGTAGT ACCAGAATTG TTTTGTAAAT GGAAAGTTTA	1560
	AGTAAATAT ATTGTAATAA AAAGGTAGAT AAACCATTGT ACAACAGTAT TCTAGGCCGC	1620
	CAACAAAAGT GTGACAGACA CACTAAAAGC CCTCCAACTT TAACTTGTA GGTAGCTTCA	1680
10	TTCTCAAAAG TGACTCCTTT TTTTCTTTT TCCTTTTCTT GAGTGTAGTA CAGTTAAAAT	1740
	TTCAAAACAG TCCTTGACAC TGCCTTTCAT GTTCAAACCA GCCATTTTGT TGTACTTTGG	1800
15	TAAAGGACCT CTTCCTCTTC CTGCTCTACA CATAACAGATA CACCCACACA CAGACTGACT	1860
	CTCTTTCTCT CATAACCCAA GGTGATGAGT GAATGATGCT TAGTTCCTTG TAAAGAAAAT	1920
	CTTGGGATGG GGAAAGGGT AGGCAGCAAG AGGATTCAAC AAACGAAAAA CATAAAAACT	1980
20	TTGTATATGA CTTTAAAAAC AAGAGSACAA CACAGTATTT TTCAAAATIG TATATAGCGC	2040
	ATATGCATGG ACAAAGCAAG CGTGGCACGT GTTTCATATA TGTTTAATTA CAAAAAATA	2100
25	TTTATCTCTT AAAAATCTTC AAGATTATGT CTATTGCTG TGCATTTTCT TTCAGTTTGC	2160
	TTATCTTTTC CGGGTTGGGG TTGGGATAAA GGTGTGTGGG TTTAGCACCT CTGGAAGACC	2220
	TATCTAGAGC TCTTTCACTT TCCTGAGGTT ATTTTGCCCY TTCTGGTGTI GGTATGTCTG	2280
30	TTGCGGCGCA TGGGCTTCAY GCCTTGAATT CCTGCTCTTG ATCAGGGACA AGGAGGTCA	2340
	AGCTCTGACT AATGCCATGA CCTGATTAAG GGGTACAGCA GGGAGTTTIG TTGCTACAGC	2400
35	TCATGAATTA ACCTGTCCCA ACCTAATCCC CCTCCATGGC ATCATGCCTC TACCCAAGCC	2460
	TTTGTGTGEC CATGTTATGC ACACAGCTGT AGGCATTCTT AAGTCCCCCT TCGCATCCAG	2520
	TGGAAGCATT TTAAAAATTC TTTACTTTT TGGTTTTCCC TTAATTGCTG CTTTTCAGAT	2580
40	TTTAGTTATG GCTCGTCTGC TCACCCCTTC TCTACATTAG GGTGTCAAAG AGAATGTTTT	2640
	GCTTTAAATA TAAATAGCCA TTCATTTAGT CTCAGATTGT GAATTTAAAA TGGTGGATAC	2700
45	CGAAATTGCT TGTGTGTGTT GCTGTGGGTT TGGTTTGAAG GCAAACACCC CTAGAACATG	2760
	ATATTCCCAT CTAGTGCATT TAAATAGAAA TCACTGAGTT TGCTGCTTTT TTATTGTCAG	2820
	CAGATAGGAG AATTAATAAT GCATTTTAGC TGTGATGTCC ATTTTATGA AATTCCTACT	2880
50	AAGAGCTATG TTAAGAGTAA AGGATGGTGG TGGTTGTATT AACTATATAC CTGTTTAGGC	2940
	CATTCTGGCT GTGGTATTTT TCAATAGGTC AGCATCTGTA AATCTGTCAG TTTTATACAG	3000
55	GAGTGCAGAG TGAAGTAGGC AACTAGATTA AGAGGTCTAA ATATGAAATA CCACTTGAGG	3060
	CTGAGGACCT CTTCGTCTTC CTTTAAATGT CTTTGCCTA GGGAGTGTIT ACCATTTGTG	3120
	AGGCAGCTTT GTCTGCTCTT AACTGTACA TCCTATTACT CCATTGGGAA GTAGGTTTAC	3180
60	TTCTCTCTGG CCTTTTGCCT AAGTTAGGCT TTGCTGAATC AACCCTACTT TTCTTTTAG	3240

AAAAGGTTBT TACAGGASAT TTACTGSCAA CTGTTCTTTT CCGATCAAAA ATCAGTGAAT 3300
 5 GTTTCCTGAG TATAAATGCT GCTTCCTTAA ACCACTTGTC GCTTTAGGAT CAACTTTAUC 3360
 TGTACCTTTT CTCTTTTCTT CCGTTGCCAC CTCAGGTGCA AATCTGAACT CAGTGTCTGC 3420
 TTCTTCCATT TTCTGTTCTC TCTCCCTCTT TCCGCCATTA TCCATATGAC ATTATTTTAC 3480
 10 TTCAAATGAC AGCATCAATC TTAAAAAGAT ATACATTAAA ACTAAGGAGT TTTTTTAAAG 3540
 AAAGCCTGAA TAAGTTCTTT TCCCTGGTAA CTTCGAAAAG CAGTCAGAGT TGCTATATAG 3600
 15 ATATATGTGG CTCTTTTAAA ATGCTTTGTG TATGTGTGGT GTTTAAAAAA AAAAAAAAAA 3660
 TTCGGGGGGG GGGCCGCTTC CCAT 3684

20

(2) INFORMATION FOR SEQ ID NO: 116:

(A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1965 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

30 AAGAAAGGCT ATTAAAAATC TAGATCACAT ATGGACCCCG GAAGGTTTTT NACCCCTCTCT 60
 TAGTGACATC GAGTCTCCCA CTAGACAAAA TAGGTGGAAA AATTTCTCGA GGGCTCACAT 120
 35 TGTTTTGTCA TCTTTAGSAA AAACACCACC AGGCCATACC ACAGCCTGCC CAGTGAGGCG 180
 GTCTTTGACA ACAGCCTCGG GATGCTGGTG GTGGCCTTTG GGCTGCTGGT GCTCTACATC 240
 40 CTTCTGGGTT CATCTTGSAA GCGCCAGAG CCGGGGATCC TGACCGACAG ACAGCCCTTG 300
 CTGCATGATS BGGAGTGAAG CAGCAGGAAG GGGCTCCCAA GAGTCCCTGS TGGTGCAGCC 360
 TGTGCTCCCC TCAGAAGGTC TCTCTTCCC AGGCTCCCC GCTGCTTTCA GAGGCTACT 420
 45 TTCTTCAAT GATGAGGACA GACTCTTSC CTGGTCTGS GCTGCCCC TCAGGCTGCG 480
 TTGCTGCTG TCTGCTTTC TTGGTGGTTC TGGCTGCTG TGGCTGCTG CTCTGCTGCT 540
 50 GCTTGTGCT TGTCTCTTT CTTTGGTGG TTTGCTGGG GCTGCTGCT CTTCTCTGCT 600
 CTGCTTGTCT CTTGTCTGCT TTCTTGTGT GCTTTGCTT CTGCTCTCT TGGCTGCTG 660
 TCTCAGGCTC TCCATTGACA CGAGGTCTC CTCCTCTGCT CCGCTCTGCT TGGCTCTGCT 720

60

AAGGCTGCTA TTTTAAATC GATGAGGCTG CAGAGGCTG GCTTCTCTG CTCTGCTGCT 780

CAGCCCCATC TGGATGTGAG GTGGGGTGGG GACATCATGG GGTGATTGCA GAAAGGGGGA 960
 GTGGGGGGGG ATGAGGTTTC TGGTGAAGAG CTGAGGGGTC TGAGGTGTTG TGTTCGAT 1020
 5 TGCTGCTCAG TGTCTGCATG TATTGTGACC GTGGGGGCTG AACTCTTCCA GCTGCTGCTA 1080
 CAGCTGAGGC CTGGATCCCG GCTTTCCCT GTGACTTACG TGTGTGTAC CCGCAGCAG 1140
 10 CCGTACAAAT CTTGGTGACC TGTCTCCCA AGAAGAGAGC CTGTCCCGAG ATGTCCAGT 1200
 AGCGATGAGT AACAGAGTG GCTGTGACT TCTCTACTT CTCTTGGTG GATCAGGGGC 1260
 TTCTGCTC CCGCTGGGCA GTTCTGGCT TGTCTCTTG GCAAGGGGCG ACCCTCTG 1320
 15 ACGCTCTGC AGCTACCAT GGAGTGAAG CCAAGTTGT GGTGTCCAGT GTGAGCAGC 1380
 CCGGGAGGC ACTGGCAGT TTAGAGGGGT TCTTGTGA GAGCCACAT GTTCACTG 1440
 20 GCGCAACAT GCTGCTTGC CTGGCCCAAC CTAGGTTGT GTGCTATCT AGAGTTGAG 1500
 CTCTGCTGT TCTCAGGG AGAAATAGG GTGAGAGG GGAAGGTTT TCTCTAAG 1560
 TGTGCTGCT GTGCTTTT TCTTCTCC AAAGAAGCAG TGGCAGGTC CAAGTTGAG 1620
 25 ACTGCTGTG TTAGTAAGCA AGTGAAGC CTGGGGTTT GAGCCAGCT ATTCTGTGG 1680
 AGCATCAGCA TCTACTCT GGAACATCA GGCACAGCT CAGCCAGGC TCACATGCG 1740
 30 AGATTTGCT AAGAGGCTA ATATTGACC TCTTACTGG CTGGAGGCTT CAAAGTACT 1800
 GGGATGTCT CAGGCACTT GGTCCCATG ACCAGTCCC CTCTCCATA GGGTAGGCA 1860
 TTCTACTGT TTATGAAGT CGAGTTTCAT TAAATATTT AAGAAACAAA GCTGTCTTG 1920
 35 TTAGGCTGC TATAACAAA ATATAATAG CTGGGTGGT TAAW 1960

40 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 503 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

50 AGTGATCCC TTGCTCGGC CTCCCAAAAT GCTGGAATTG TAAGCGTGG CCTCTGACC 60
 CGGCTGCTC CGCAATTAA AAAGGCACAG CCAACATTCC CT/TCCAGAA AGCAGCAGA 120
 TGCTTTGGG AGAACCAGCC TCTTCATGG AGGAAAGCTT GGGATCTCC TTCCACCTG 180
 55 GGGAGGAGAG GATCTGTGG AAAATCCTTC TGACGGACTT CCGCTCAGTG CTTGATCCAT 240
 ACTCAATAGT AGAAAAAGTA AGAAATATAC AAAGATAGCA GATACACGA GACAGTTCCC 300
 60 CAAATAGCTG AGCGAWTAGC GCAGAAGCAA TATTGAAGAC CTAATAGTTG AGACATTTCC 360

371

AGAACTGATA AAGTGCATCC AGCCACAGAT CAAGCAGCCC AGAAAATTCG AGGCAGCATE 420
 AACAAATAAA TAGCCCCACA TGCACCCGTG AAAATGCAGA AGACCAAAAC AAAAAGTCOS 480
 5 GTCACACAGCC AGAGTTAAAG AGG 503

10

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1133 base pairs
 15 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

20 GGCACAGCTT GGAATGAACC CCTCTGGATA AGGGGGACTA TTAGATAGAA TAAACATCAA 60
 TAAATGCTTG ATGAATAAAC GCTAATCCTA CCTTCCCAGC CTGACACCTC CCAGTGGACA 120
 25 CCACACTTCA CTTGAAGCCT TAGAAACCTT TCCCACCCAT GCTTCCAGCC CTGGCTTCAT 180
 GTTGCATTTT CTCACCCCCA GAACAGGCGG CCGGCTGAA GAAACTACAA GAGCAAGAGA 240
 AACACAGAA AGTGGAGTTT CTTAAAAGGA TGGAGAAAGG GTGTCCAGAT TTTATTCAAG 300
 30 ACAGTGGGCA GATCAAGAAA AAGTTTCAGC CAATGAACAA GATGAGAGG AGCATACTAC 360
 ATGATCTGCT GGAAGTGGCT GGGCTGACAT CTTCTCTCTT TGGGGAAGAT GATGACTGTC 420
 35 GGTATCTCAT GATCTTCAAA AAGGAGTTTG CACCTTCAGA TGAAGAGCTA GACTCTTACC 480
 GTCTGGAGA GGAATGGGAC CCCCAGAAAG CTGAGGAGAA GCGGAACNTG AAGGAGTTGG 540
 CCGAGAGGCA AAGAGGAGGA GGCAGCCAG CAGGGGCTTG TGGTGGTGA GCTTGGCAGC 600
 40 GACTACAAAG ACAAGTACAG CCACCTCATC GGCAGGGAG CAGCCAAAGA CCGAGCCAC 660
 ATGCTACAGG CCAATAAGAG CTAAGGCTGT KTGCCCTGG CCAATAAGAG GCACACACCC 720
 45 TTTATTCAAG AGGTATGAA TAAATGAGA GCGAAGAAC GTCTGGGA GATGGGGAA 780
 GAGTGGGGG CAAGTCTTA GGGGGGGG CAGGTCTCT TGAACCTTG GGGCAGGTA 840
 GGGGGCAGGG AGAGCAAGG CTCTGCTAT TAGAGCCAT CTTGAGGCC CAGCTCTGAA 900
 50 CCAGTCTTA CCAGTCTCC CTCAGCTGG GGGAAAAG GTGTTTGAAT TGTCACTTT 960
 CCAGCTTGA TATGTGGTG GATGTGTGT GTCTGTGTA GAGTGTGAAT CCACAGGTGG 1020

60

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1101 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

```

5      GGGCACAGCT GAAGCTGCAG ACCTCCCCAG GGGATGKCTC CTCTCCCCCA GGAGCCCCGA      60
      GGCAGGGGAG GCAGAAAGCC TGGGTCTCTG GGGTGGGCT GCGGACAGCT GTGCTGTGGG      120
15     GCGGGGGGCTG GGCCTGTCCC ACAGGAGGCT GGAGCTGTG GTTCTGAGCA GCGAGCTGGG      180
      TGGTGTCTGG GATACCTGG GAGGACAGC GGTGGGATG TGGGACTGGG ACTGGAGTGG      240
20     TGGTGGTCT TGGCTGTCT GGTGAGCTT TGTGTGTGTC TGGCTGAGTG CAGGGGGCCAA      300
      GGGGCACAGG GGTAGTAGG GGGGACAGC TGGGGGCTC ACCTGTGAGA TGGGGTCGGA      360
25     ATTTACACA GGTATAGCT TGGTCTCTG TGTGAGAGC TGGACTYCTK ASAACGGGAG      420
      TGTGTGTCT GAAAGGGCTG GTTGGAGAGC AGCTGCTTTT CTCGCTGTTC TTCTCTTAGG      480
      AATTAACA AAAACAGAAA GCACAGAGC AACTCAGTAG CAGACCCAG AATCTCCCT      540
30     TGGCAGAGT GGTTCAGAC GGGGAGAGC AACTGTCCA GAAGGGATT CAGCTGCTCA      600
      AGGGGCTTC GCGGGGGGCT GTCCAAAAC TGGCAGGGCT CCAGTAGGCC AATGGGCACC      660
35     AATGAGTCTT GGGTGGGGCT CTGGGAACT TTTTGTGAT AGTTGGGTTT GAGGCTTTG      720
      CTTACAGGT CAAGTACGT CTGAGGAGCA TGGGCAGGA GTGAGGCCCA GTGGCCGAGA      780
      CCAAGGGG CACTGAGGT ACAGGAGC AGAGCGTAC CTCGGCAGGC TGGACACACT      840
40     GCCAGCACA GGCAGAGCA CCAGGTCTT AGSTTTAGCT TTTAAAAACC TGAAGGGGA      900
      AGCAAAAACC AAAATGTGT ACTGGGCTTT GGAGGAGACT GGAGCCTCAG CCTGTCTCTG      960
45     GCCACGGGCC GCTGGGGCTG GTGTGGGTG GCTTGTGTG CTGGATTGT AGCTTATCTT      1020
      CCGTGTGTG TTTGAGCTG TTTTAGTAAA CCGTTTTTTC ATTTTAAAAA AAAAAAAAAA      1080
50     AAATTTGGG GGGGGGCCCC N      1101

```

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 282 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AGCTTCTCTG TCCAGTCTTG AACTCTGGGS TCTCTTGGAA CTTTCTCAGC CCTCTCAGC 60
 5 CTGAATATTC CTTTCATGGA TTCCACTCAA CCAGACTTTG CATGTGTGCC TACTTAATCA 120
 ACCTTATCTT TGAATATST TCGGCCCCAC CTTCCACTCC TTGTTCTTIG TTCTCTCTTG 180
 10 GCGTAACCTG TGGTTCTTC AATTCCATC CCGCGTGGGA CAGCATTCCT CCTTCTCTCC 240
 AACCTCCCTC CGTCTCARAA AAAAAAAAAA AAAAAAAAAA TT 282

15

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 2635 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

25 TAAGGGGCTG TGTCTCACC TCTCTCTGAC CCTTAACACT CCGTCTCTCC CCAGACCAAC 60
 AGAGAGAGCT GTTCTCTAGA CCGCGGAGAG AAGCAGCTGC CGAAAGCTGC AGCTTTTCCG 120
 30 CACTCTGAGA CATGATCTT CTTCTCTGCA GGGGAGAGCC ACCACAGGC CATCTCAGC 180
 CCACTTCTCC TCAGTCCCA GGGTTCTCTT CTGGCTCTCC TGAGGATTC CTAGGCTCTC 240
 35 CCGGAGAGG GGYTTCCCA AGTCTCTTT TGAAGCTGC AATGTGAAA AGTGAAGT 300
 CAGAGGGAAC AGGACAGCTC CAGCTCTCTT CTGAGGCCAC AACTCAGCC TGGTCTCTCC 360
 CCAACATCCC CTGACCACTG TGAGTCTATC TCACAGATG AGAAGAGGCT CTGTCTCTTT 420
 40 YTTTTCTTTG TTTCTTCTG TTTTCCCA CCACTCAGT TCTCTCAGC AAAACAAAT 480
 CCTTAACATC TTCTCTGAG AATTCTCTAC CCAGACTTGC GCTCTCTATC CCTTCTCTG 540
 45 CCGTCTGAGT CCAGCTCTG TGGTCTCTG TGTCTCTGA CCGTCTCTGC ATCTCTCTG 600
 CCGTCTGAGT TGAGGAGAG CCGCTCTCTT CCGTCTCTG TGGTCTCTG GGGTCTCTG 660
 CAGTCTGAGT CTCTCTCTG GGTCTCTCA CCGTCTCTG CCAGCTCTCC AGCATCTCTG 720
 50 AGCTCTCTGT GCACTTTTCA GAAGAGCTCT ATCTCTCTG CCGCTCTCTG TGGCTCTGAA 780
 TCAACATCTT CCGAGCTCTT CTTCTCTGAA ATAGAGAGT CCACTCTAAC TGCATAAAT 840
 CCGAGCTCTT TGAAGCTCT CCGTCTCTG CCACTCTGAG CTAGCTCTCT CTTCTCTCTG 900
 60 CCGAGCTCTT TGAAGCTCT CCGTCTCTG CCACTCTGAG CTAGCTCTCT CTTCTCTCTG 960

	TCCCAGCTGG AGTCTCTGGAA CTTTCTCTCT CCGGCTGGGG GTGGGGCTTG TTAAGGATCG	1140
5	TGGGGGGGCTT GGGGAAGGAA GGAGTTCAGA GGAGGCTGT CCGCTCTCTT CTTTATGTCA	1200
	CCCTCTCTTC CTGGGACACG TCTCTCTCTT GTCTCTCTCT CTTCTGCTTG TCACTCTTG	1260
	TGTGTCTCTG TAAATATGTT TTAGGAAGAA AGCAAAAGAG ACTGAAGTAG CTTCTGCTAG	1320
10	GATTGAGGG GTCCAGCCTT GCCTCTCTTC GAAGCTCTCA CACTCTCTTT CCGCTCTCTG	1380
	AGACTCTCTC CTTCAAAAGG TAGACAAAG AGCAGCTCTC TGTGAGCTTG AAGGGGGGGG	1440
15	TCAAAAGTGGG TTTTCTCTAG ACAAGCTTAA GCTTCTCTCA TGAGCAAGCT TCTGATCTGG	1500
	TCTTCTCTCA GCTCTCTGAT TTGTGACCTT GACCAAGGGG CCTGCTCTCT ACGCTCTCTA	1560
	GTGCTCTCTC CTGATCTCTT CCGCTCTCTC TCGCTCTCTT CCGCTCTCTT AAGTCTCTAG	1620
20	GGCAATTAGG GCACTCTCTG AAGAAGCTTA ACCCTCTCTT GAAAGAGAGG TTTCTCTCTT	1680
	GCTTCTCTCT GCACTCTCTT TCGCTCTCTT CAGCTCTCTT TCGCTCTCTT GTCTCTCTGG	1740
25	AGCTCTCTCT CAGCTCTCTT AGCTCTCTCT AGCTCTCTCT AATCTCTCTT GAGCTCTCTT	1800
	TCAGCTCTCT TCAAGCTCTG ATGAGAGAGG AAGCTCTCTT TGGGAGCTTG GAGCTCTCTG	1860
	GTCTCTCTCT GCGCTCTCTT GCTCTCTCTT TCTCTCTCTA CAGCTCTCTG GTCTCTCTCT	1920
30	CTCTCTCTCT CAGCTCTCTT GCTCTCTCTT TCTCTCTCTT TCTCTCTCTT TCTCTCTCTA	1980
	CTCTCTCTCT GCTCTCTCTT GTCTCTCTCT ACCTCTCTCT CCTCTCTCTT ACCTCTCTCT	2040
35	GAGCTCTCTG CCGCTCTCTG GAGCTCTCTG CTCTCTCTCT ATCTCTCTCT TCTCTCTCTG	2100
	GCTCTCTCTT GCGCTCTCTT CAGCTCTCTG GAGCTCTCTG AAGCTCTCTT AAGCTCTCTT	2160
	AGGCTCTCTT CAGCTCTCTG AACTCTCTCT CAGCTCTCTG CCTCTCTCTT CTTCTCTCTG	2220
40	TGCAATAACT GAGCTCTCTG TTAGCTCTCT TGTCTCTCTT CTTCTCTCTT CTTCTCTCTG	2280
	TTTAACTCTT GAACTCTCTT CTCTCTCTT TTTCTCTCTT CTTCTCTCTT ATTCTCTCTT	2340
45	AAAGTCTCTG ACTCTCTCTG TGTCTCTCTG TGGCTCTCTA AGCTCTCTCT GATCTCTCTA	2400
	GCTCTCTCTT TCTCTCTCTG CCGCTCTCTT TCTCTCTCTT CCGCTCTCTT ACTCTCTCTG	2460
	CCTCTCTCTT GTCTCTCTTA TCTCTCTCTT ACTCTCTCTT AGCTCTCTCT TACTCTCTCT	2520
50	ACATGCTCTA AGCTCTCTCT ATCTCTCTT TAAGCTCTCT GAGCTCTCTT ACTCTCTCTT	2580
	CACCTCTCTT AACTCTCTCTT CCGCTCTCTT GCGCTCTCTT ACCCTCTCTT CTTCTCTCTT	2635

55

(2) INFORMATION FOR SEQ ID NO: 122:

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 994 base pairs

375

(E) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

	GAATTGGGCA GAGGTTCGGG GAAATAGGG AATAAGGAAG CACAGGAGTA GGGGAGAAGG	60
	AAGCACAGGA GTAGGGGAGA TATACAGGG TTAGGATAAG GGGGAAAGGG CGGTGGTTTC	120
10	SCAAGAGSTG AAACAAGATG TGAGAGACAA GGGGTAGGSA AGAAATGGGG CACGGGTAG	180
	GITCAGAAAG CCATAGAGGG TGCGGACGG GGAATGCGAG GGGCACAGAA AGGAATGAG	240
15	GGGTGGGCTA TTTTAAGGA GATGGTCTT CAGGCTCTT YTTTCTGCG TATTTCTCT	300
	CTTCAGGCG GCGGCGGGAT ATGTCTTGG GAAACCGGC CATTGTAGG TGGATGATGA	360
20	GGCAGCTCT TCTAGGTTC TCAAAGACTA CCAGAATGTC CTTGGAATG AGAAGGTGA	420
	TCATGTCTG AAAAGACTCT TGTCTTTGA AATGGCCAAC AAGAAGGAGA TGCTAAAAAT	480
	CAAGCAAGAA CATTATTGA AGAAGATTGT TGCAAAAGCA GAGGACACCA GATGCTTGA	540
25	GGCTCGAATT ATTGGCTTGT CTGTCAAGAT CGGCAGTTAT GAAGAACACT TGGAGAAACA	600
	TGAAAAGGAC AAAGGCGACA AACCTATCT GTTAATGAGC ATTGACCAGA GGAAGAGAT	660
30	GCTCAAAAAC CTGGTAACA CCACTATGA TGTCTTTGAG AAGATATGCT GGGGGCTGG	720
	AATTGAGTAC ACCTTCTTCC CTCTGTATTA CCGAAGAGCC CACCGCGGAT TGGTACCAA	780
	GAAGGTCTG TGCATTCGGG TTTTTCAGGA GACTCAAAAG CTGAAGAAGC GAAGAAGAC	840
35	CTTAAAGGCT GCAGCAGAG CCCAAAAACA AGCAAGCGG AGGAACCCAG ACAGGCTCTC	900
	CAAAGGCATA CCAAGACAC TCAAAGACAG CCAATAAAAT CTGTCAATC ATTTAAAAAA	960
40	AAAAAAAAA AAAAAAAAAA AAAAAGGGGA GGGG	994

45 (2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1542 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

60 GGGTTCGGG TATTAATGAG GGGGAGGCG GGGGAGCTG TGGAGAGT AGAGGAGT

	CTGGACATCT GATGAAACAG TGGTGGCTGG TGSCACCGTG GTGCTCAAGT GCCAAGTGAA	300
5	AGATCACSAG GACTCATCCC TGCAATGGTC TTAACCTGTC TCAGCAGACT CTCTACTTTG	360
	GGGAGAAGAG AGCCCTTGGG GATAATCGAA TTCAGCTGGT TAMCTTACG CCCCACGAGC	420
	TCAGCATCAG CATCAGCAAT GTGGCCCTGG CAGAGAGGG CGAGTACACC TGCTCAATCT	480
10	TCCTATGCTT TGTGGGAAGT GCCAAGTCCC TGCTCACTGT GCTAGGAATT CCACAGAAGC	540
	GCATCATCAG TGGTTATAAA TCTTCATTAC GGGAAAAAGA CACAGCCACC CTAAACTGTC	600
15	ACTCTTCCTG GAGCAAGGCT GAGGCCCGGT TCACCTGGAG AAAGGGTGAC CAAGAAGTCC	660
	ATGGAGAACC AACCCGCATA CAGGAAGATC CCAATGGTAA AACCTTCACT CTCAGCAGCT	720
	CGGTGACATT CCAGGTTACC CCGGAGGATG ATGGGGCGAG CATCGTGTGC TGTGTGAACC	780
20	ATGAATCTCT AAAGGGAGCT GAGAGATCCA GCTCTCAAGC CATTGAAGTT TTATACACAC	840
	CAACTGGGAT GATTAGGACA GATCCCTCCC ATCCTCGTGA GGGCCAGAA GCTTTGCTAC	900
25	ACTGTGAGGG TGGGGGAAT CCAGTCCCGC AGCAGTACCT ATGGGAGAA GAGGGCAGTG	960
	TCCACCCCTT GAAGATGAGC CAGGAGAGTG CCTGTATCTT CCTTTCTCTT AACAAGAGTG	1020
	ACATGGCAC CTATGGCTGC ACAGCCACCA GCAACATGG CAGCTACAAG CCTACTACA	1080
30	CCTCAATCTT TAATGACCCC AGTCCGGTGC CTCCTCTCTC CAGCACCTAC CAGGCCATCA	1140
	TGGTGGGAT CTTGGCTTTT ATTGTCTTCC TGCTGCTCAT CATGCTCATC TTCTTGGCC	1200
35	ACTACTTGAT CCGGCACAAA GGAACCTACC TGACACATGA GGCAAAAGG TCCACGATG	1260
	CTCCAGAGG CGACACGGCC ATCATCAATG CAGAAGGCGG CCAGTCAGGA GGGACGACA	1320
	AGAAGGAATA TTTCATCTAG AGCGGCTGCG CCACTTCTTG GCGCCGCCAG CCGCTGTGG	1380
40	GGACTTGTG GGGCCGTCAC CAACCCGGAC TTGTACAGAG CAACCCGAGG GCGCGSCCT	1440
	CTCGNTTCTT CCCCAGCCCA CCCACCCCTT TGTTACAGAA TGTYTKGTTT GGGTGCGGT	1500
45	TTTGTWATG GTTTGGGATN GGGGAAGGGA GGGANGGCGG GG	1542

50 (L) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1390 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

60

CAAGCTCTAA TACGACTCAC TATAGGGAAA GCTGGTACGC CTGCAGGTAC CCGTCCGGAA 60

377

	TTCCCGGCTC GAGCCAGCTG TCGGCGGCTC AGGGTGGAGG CATGGTTCTG CACTGAGGCC	120
	CTGTGATG TGCGGCTGT GTGCTACTTG GTAGCGGCGG CTCTGCTAGT CGGCTTTATC	180
5	CTCTTCTGTA CTGTCAGCGG GGGCGCGGCG GCATTCAGCGG GCGAAGAGGC ACTGCACAA	240
	GAGGAGCTGG GAGGAGGAGG CGGCGTGGGC CAGGCTGGGC CCGTGGAGGC TGGGAGCGG	300
10	AGAGCTGAGG GAGGCTGCG GGGCGGAGG GAGCTGGGCA GCGGCTTACA GCGGAGCGT	360
	CGAGCGGAGG GGGTGGCTTG GGCAGAGCA GATGAGAAGG AGGAGGAAAG TGTGATCTTA	420
	CGCCAGGAGG AGGAAAGTGT CGAGAAGCTA GCGGAAATTC AGCTGTGGG GAAATTTGGA	480
15	GCTAAGAAAC TCGGAAANN GAGGAGAGAA CAAGCGGGA AGGCGCAGG TGGGCGAGG	540
	GAGGCTGAAC GTGAGGAGG GAAAGGACTC GAGTCCAGG CGGATGAGT GGAAGAGGA	600
20	GGAGGAGCG CTTCGCTGAG AGGAGGAGCA GAAGGAGGAG GAGGAGAGGA AGCGCGGGA	660
	GGAGGAGCG GAGGCGGAG ATGAGGAGTA CCGGAACTG AAGGAGCGCT TTGTGCTGA	720
	GGAGGAGCG GTAGGAGAGA CCGTACTGTA GGAACAGTCC CAGAGTTTCC TGACAGAGTT	780
25	CATCAACTAC ATCAAGGAGT CGAAGGTTGT GCTCTTGGAA GAGCTGGCTT CCGAGGTGG	840
	CGTACGAGT CAGGACAGCA TAAATGGCAT CAGGAGCTG CTGGCTGAGG GAGCTATAAC	900
30	AGCTGTGATT GAGGAGCGG GCAAGTTGAT CTACATAACC CGAGAGGAG TGGCGGCTT	960
	GGCGAACTTC ATCGAGAGG GGGCGCGGCT GTGATCGGC GAGCTTGGC AAGCGAGCA	1020
	CTGCTGATC GCTGGGCGG GGGAGTGGGC TGGCGAAGGC CAGGCTGAG CCGAGTCTT	1080
35	CGCTCTTGA CTCAGAGTGG GTGTGGCTTA CCGGCTATA CATCTTCAT CCGCGGAGC	1140
	ATCTTGGGGA AGTATGCTG TGGCTAGGCA GTTATAGATT AAAGGCTGT GAGTACTGT	1200
40	GAGCTTGGTG TGGTTGGTG TGGCAGAAGG CCGGCGTAG GATCTTAGAT AAGCAGGTGA	1260
	AATTTAGGCT TCAGATATA TCGGAGAGGT GGGGAGGCT CCGTGGAGC TGGTGAAGTC	1320
	CTGTTCTTAT TATGAATCCA TTCATTCAAG AAATAGGCT GTTCAAAAA AAAAAAAAA	1380
45	AAAAACTGGA	1390

50 (2) INFORMATION FOR SEQ ID NO: 125:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1388 base pairs

60 GAGGAGCTGG GAGGAGGAGG CGGCGTGGGC CAGGCTGGGC CCGTGGAGGC TGGGAGCGG

5 GACGCTGACC AGCTTCCCTCT CCTCGGTCTC CTCCGGCTCC AGCTCCGGGC TGCCCGGCAG 120
 CCGGGAGCCA TGGACCCCA GGGCCCGCCC GGTCCCGCGC AGCGGCTCCG CGGCTCCTG 180
 CTGCTCCTTC TGCTGCAGCT GCGCGCGCCG TCGAGCGCTT CTGAJATCCC CAAGGGGAAG 240
 CAAAAGGGAJ ATCCGGCAGA GGGAGGTGGT GGACCTGTAT AATGGAATGT GCTTACAAG 300
 10 GGCAGCAGGA GTGCTTGGTC GAGACGGGAG CATTGGGBCJ AATGACATTC CGGTACACC 360
 TGGGATCCCA GGTGGGATG GATTCAAAAG AGAAAAGGGG GAATGTCTGA GGGAAAGCTT 420
 TTAGGAGTTC TGGACACCCA ACTACAAGCA GTGTTCATGS AGTTTATTGA ATTATGGCAT 480
 15 AGATCTTGGG AAAATTGGGG AGTGTACATT TACAAAGATG CGTTCAAATA GTGCTCTAAG 540
 AGTTTTGTTC AGTGGCTCAC TTGGCTAAA ATGCAGAAAT GCATGTGTTC AGGTTGGTA 600
 20 TTTCACATTC AATGAGCTG AATGTTCAGG ACCTCTTCCC ATTGAAGCTA TAATTTATTT 660
 GSACCAAGSA AGCCCTGAAA TGAATTCAAC AATTAATATT CATGGCACTT CTTCTGTGGA 720
 AGGACTTGT GAAGSAATTS GTGTGSAAT AGTGGATGTT CTTATCTGGG TTGGCACTTG 780
 25 TTGAGATTAC CCAAAAGGAG ATGTTTCTAC TGGATGGAAT TCAGTTTCTC GCATCATTAT 840
 TGAAGAACTA CCAAAATAAA TGCTTTAATT TTCAATTTGCT ACCTCTTTTT TTATTATGCC 900
 30 TTGGAATGCT TCACTTAAAT GACATTTTAA ATAAGTTTAT GTATACATCT GAATGAAAAG 960
 CAAAGCTAAA TATSTTTACA GACCAAAGTG TGATTTTACA TGTTTTTAAA TCTAGCATT 1020
 TTCAATTTTC TTCAATCAAA AGTGGTTTCA ATATTTTTTT TAGTTGGTTA GAATACTTTC 1080
 35 TTCATAGTCA CATTCTCTCA ACCTATAATT TGGGAATATT GTTGTGGTCT TTTGTTTTTT 1140
 CTCTTAGTAT AGCATTTTTA AAAAAATATA AAAGCTACCA ATCTTTGTAC AATTTGTAAA 1200
 40 TGTTAAGAAT TTTTTTTATA TCTGTTAAAT AAAAATTATT TCCMACAACC TTAACAAAAA 1260
 AAAAAAAAAA AAAAAAAAAA AAAAAAAA 1288

45

(2) INFORMATION FOR SEQ ID NO: 126:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1517 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

AGTGGCTTAA AGGCATCGTT TTAGGGATTA CTGGGAAGTA TCTTCAAAGT AATACATGAG 60
 AAACATTCCT TCCTAAATCC TTTATTATAT TGAATATCGT ATTAATTGGT TTTCAGAGGT 120
 60

TAAATTAAACC ATGTATTTCCT GCAATAAAATG TCACTTGTINT CTTGTATATA ATCTTTTTTA 180
 TATATTAGCG GATTGATTCA TTAGTATTTT GTTGAGGATT TTTGTGTETA TATTCATAAG 240
 5 AGATGCTGCT CTGCAGTTTT CTTTTTTTTT GATAATCTGG TTTTGTATC AGTAATACAG 300
 GCCCCATGAA ACGAGTTGGG AAGTGTTCAC CTCTCTTGTA TTTTTCAG AGTTTGTGAA 360
 GAATTCCTAT TAATTCCTTA AATGTTTGTG AGAATCTACC ATTGAAATCA TGTGTCTCTG 420
 10 GCTTTTTTTT GAGGGAAGTG TTCTGATAAC TAATTCAGTA TCTACTTTTT ATACCTCTGT 480
 TCAGATTTTG CTCTTCTCTG AGTTAGTTTT GCTAATTTGT GTATCTCTAG GATTTGTGCC 540
 15 ATTTTCATTA TCTCATTTGT TGSCATAAAT TAAACTAAAT TTGGCCTGAG CCTACCTGTA 600
 TATCTTGAGT CCTCTGTAA GGAAGTGTAG CCTAAGTTGT ACATAAAGAA ACTGAAATCC 660
 TAAATTAGGA ATGTAGTTTT TSTAACAGCT CCTGAGTCTC AGGCAGTCA AGCAGYCAAG 720
 20 TCTGTCAATT GCAGGCTGCT AACTAAGCAG CCCATGSTCA AATGAGGCAA AACCTTTTCC 780
 TTTTAACACA TAGTATAGCT TTGTAATCCT TTTCTTGCAC ACTCGGGTAA TTTCTTCTT 840
 25 TTTCAATCCC KGWATTTTCC AKGAATATGA RTCTYCTTT TTTCCCTCC TGTGAGTCTA 900
 GCTAATGCTT TGTCAATTTT GTTGATCTTT TGAARAACAA ACCTTTGGTT CCACTTTCTT 960
 GTTGCATATG CTGATATTC TCATAATTGG AGTGGAAAGC TGATCTTTGA TTAATTATTT 1020
 30 TACTTAGGCG TGAGGAGTTC ATGGACTTGG CAAAACCTCC TTGAATCTAA ATTGCATCTT 1080
 CTTTCTCTGT TTCTGGGCTG AAACATGTTT TTTCCCATCT WANAWACCT TGGTCTTTTC 1140
 35 ATGAGGAGAT AAGACTAGAG AAAGTTCTAG ATMCCTTGTG CTTTTATGCT GTCATTTTGT 1200
 TTAAAGGCTT TCTATGTAGT AAAACTATCT ATATAGACAA AATAGAGCCT TGATTTGTGG 1260
 TCTTGAATTT GATCAACATG ATTTACCACA TTCTGTACTG GATATTTCTT CACCTGCTGC 1320
 40 TACTGTAAAC CATTTTATTC TTGGATCTTC TGTAGAGTAT ATTATCACAG STACTTTTTA 1380
 CAGGGCTGTC TAACTTTTTG GTTTCCTTGG GCAGATTGAA AGAAGAAGAA TTGTCTTGGG 1440
 45 GACACATCA AATAGGCTAA CACTAATAAT AGTTGATGAG CTAACAAAAA AAAAAAAG 1500
 CAAAAAAGN CCAAAA 1517

50

(2) INFORMATION FOR SEQ ID NO: 127:

(a) General Description:

(b) Specific Description:

60

(c) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

TGAATCTATT CTTTGAACAT TGTACAAACAA GAATTACATT ATACTGTTAT ACACAGAGTAC 60
 TCTTGCAGTG TGAAATAGAT TGGTTTGSA AATGAACCTG GCTTTGCTAT AAATTACATT 120
 5 CACAGGCCCTT TTTGAAAATG TGTAACTTGC CTATCAAAGT AATTGTTAGG GCAAATGCAG 180
 AATATATGTC TCCATCTGCT AAAGTACCTT WTATTCATGT GGGAAATCAA GTAGTATCAG 240
 10 AACTTGTCTC AATACTCCAA TTTGTTAAAG CCAAGGCCCA TTCTCTTAGT GATGGGCTGG 300
 AGGAAGTCCA AAAAGCAGAA AAGAAAGCTT ACATGGAATT AGTCAACAAT ATGCTGTTGA 360
 CTCACAGAGCT GATCTTTCAG TGGTGTGATG AAGCTACAGT AGGGRMGATC ACATCATMTA 420
 15 GSTATGGWTG TCTTTACCTT TGGGCTCTGW WTCATATTTT GGCCTATCAA AAACAGTGGG 480
 AAGTCAAAAG TAAGNTGAAA GCTATTGATG GGGGAAAGAA GACTCTGGAC CAGGTCTTAG 540
 20 AGGATGTAGA ACAGTGTCTT CAAGCTCTCT CTCAAAGACT GGGAACACAA CCTATTCTCT 600
 TCAATAAGCA GGTACTGAA CTTGACGCAC TGGTATTGG CCATCTATAC ACCATTCTTA 660
 CCACACAATT GACAAATGAT GAAGTTCTCT AGAAGGTGAA AACTATAGC AACCTCTTG 720
 25 CTTTCTGTAG GAGAAATGAA CAGCACTATT TTGAAGATCG TGGTAAAGGC AGCTCTGCAT 780
 AGAGTTATGT GTTAGTCTCA GAGTCTTAA CTTTTGAAAT ATGTTTTACT TGAATGTTAC 840
 30 ATTAGATATT GGTGTCAGAA TTTTAAAGC AAATTACTGC TTTTGAAC CTCAAATTAT 900
 ATAATGATC TTATGTATGT GCTTTATATT GTTATTGTG TATACATTAA AATAATTCTG 960
 AATTATTTAA TCTGATATGT TGTATTCTGT ATCTTGAAAT TTTGTTTCC TTGAAACATG 1020
 35 CATGCATTTA AAAATAAAGC TTAACAACCT GTAAAAAAA AAAAAAAAAA CTC 1073

40

(2) INFORMATION FOR SEQ ID NO: 128:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 300 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55

CAACCCCTGC CTTTTTTTG TTTTCCATTT GCTTGGTAGA TCTTCCTCCA TCCCTTTATT 60
 TTGAGCCTAT GTGTGTCTCT GCGCGTGAGA TGAGTCTCCT GAATACAGCA CACTTACTGG 120
 TCTTGAATCT GTATCCAATT TGCCAGTCTG TGTCTTTCAT TTGGAGCATT TAGCCCATTT 180
 ACATTTAAGG TKAATATTGT TATGTGTGAA TTTRATCYTR TCATTATGWT GTTAGCTGGT 240
 TATTTTGCTT GTTAGTTGAT GCAGTTTCTT CCGGCATCA ATGGTCTTTA CAANTTGGCA 300

60

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1275 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

GGCAGAGCCT GTCCCTBECTG CCCTTCGAAA AAAAACCCCC TCTGGTGTGA GCAGGATGGT 60
TGGAGGTTAT GTGAGCTCCT TCTCTTTTTC TCCAGTTTCC TCTTCCCTTC TCCTCCCTGC 120
CTCTTTTGGT TTTCCTTTTC TTCTGTGTAC CCGCTGCCCA TTCTGTATT TTCTCCATC 180
GCCATTCTCC CCTCTCCCA TGTCTCTAAC CCGTTCAAA TCTTCTCT TAAATGTTG 240
AGATTTTTC TCACCAAGCA CAGCCAGTA TTAATTAAAC TAGCTGCAA CAGGAGCAA 300
GTGGTCTACC ATGACAGATG GGTTTTGTGT GTGTGTGTGT GTGTGTAAT GTAATAAAAC 360
ATATTGARTC ACTCAATAAA CACAGAGTGT CTAATACATG TATCARGCAC TATCATAGAT 420
GCTAATTAAC GAAACTGAAA TGGCCAGGTC CTCACAGTGG CTCATGCCA TAATCCCAGC 480
ACTTTGGGAG GATGAGGCAG GAGGATCACT TGAGGCCGGG AGTTCAAGAC CAGCCTGGGC 540
AACATAGTAA GACTCCATCT CTACAAAAA AAAATTTTTT TTATTATACT TTAAGTTTTG 600
GGTTACATGT CCAGAACGTG TAGTTTTGTT ACATAGGTAT ATACGTGCCC TGGTAGTTTG 660
CTGCACCCAT CAACCCATCA CCTACATTAG GTATTTCTCC TAATGTTACC CCTCTCCTAG 720
CCCCCACCC CGTGAACAGC CCGGTGTGT GATGTTCCC TCCCTGTGTC CATGTGTTCT 780
CATTTGTTAA CTCTCACCTA TGGAGTGAGA ACATGTGGTA TTGSTTTTC TGATCTTGTG 840
ATAGCTTGGT GAGAATGTHG GTTCCAGCT TTATCCACST CCGTGCAAAG GGCATAAACT 900
CATCCCTTTT TATAGTTGUA TATGTTTGA TGGTGATAC GTGCCACAT TTCTTAATCT 960
ATCATTTGATG GACAAATTTT GTATTTGTGA ATAGTGCCAC AATAAACATA CGTGTGCGTG 1020
TGTCTTTATA GCAGCATGAT TTATAATCT TTGGGTATAT ACUCAGTAAT GGGATCACTG 1080
AGTCAAATGG TATTTCTGCT TGTAGATCG TAAGGAATTG CCACACTGTC TTCCACAATG 1140
TTTGAACATA TTACACTCC CAGCAACAT GTAAAAGTGT TTCTATTTTT CAGCAACCTC 1200

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

10 CNGAAACCCC GTGAACCTC CCGGGTTAA AAAGCCCCC CTAAATGGGG GGAACGCYTC 60
ACACGTATA AAAAAGTACT AGAATGTTTT GAAAGCGAGA AACACAGCT GTGTAGGGTA 120
15 GCTAGCAGTT AGTGTGTAC AGAAGACAGA TATTTGTGGA TTTVTGCATT TTCTAAGTTT 130
GCTGCAATGA GCATGTATTA CTTTCATAGT TATAAACAC ATGCAAAATG CCCTTTTAAA 240
ATGAAAAAAA ATCCATGAGT GTAAGTGATA TATATGCTTT GGAAAGCCTG GGACGGTCAT 300
20 TGTTFCTCT CAATAGTATG TGTTCGCTT TGTCTTTTGG AGACATTTTG TTTAATCTG 360
TTGATGACAA TAACTGTTG ATAATATAAC TTGATAACAA ATAAATGAC TTATGATTGA 420
25 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA NN 472

30 (2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS

- 35 (A) LENGTH: 1950 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

40 ACCTCTCAGA ATCTTCTCTC AGCAACCTGA GTCTTCGCCG TTCCTCAGAG CGCCTCAGTG 60
ACACCCCTGG ATCCTTCCAG TCACCTTCCC TGGAAATTCT GCTGTCCAGC TGCTCCCTGT 120
GCCGTGCCCTG TNAITCGCTG GTGTATGATG AGGAAATCAT GGCTGGCTGG GCACCTGATG 180
45 ACTCTAACCT CAACACAACC TGCCCTTCTT GCGCTGCCG CTTTNTGCCG CTGCTCAGTG 240
TCCAGACCTT TGATTCCTGG CCCAGTGTCG CCAGCCCCAA ATCTGCTGGT GTCAGTGGCA 300
50 GCAAAGATGC TCCTGTCCCT GGTGCTCTG GCCCTGTGCT CAGTGACCGA AGCTGTGCCT 360
TGCTCTGAT GAGCCCCAGC TCTGCAACGG GCACATGGGG GGAGCCTCCC GCGGGTTGA 420
GAGTGGGCA TGGGCATACC TGAACCCCT GGTGCTGCGT AAGGAGCTGG ATGGCTGCT 480
55 AGAGAACGAG GGCAGTGAGG TGCTGGCGTT GCCTGAACTG CCCTCTGCCG ACCCATCAT 540
CTTCTGGAAC CTTTGTGGT ATTTCCAACG GCTACGCTG CCCAGTATTC TACCAGGCT 600
60 GSTGCTGGCC TCCTGTGATG GGCCTTCGMA CTCCAGGCC CCATCTCCTT GGCTAACCCC 660

383

	TSATCCAGCC TCTGTTGAGG TACGGCTGTT GTGGGATGTA CTGACCCCTG ACCCAATAG	720
	CTGCCACCTT CTCTATGTGG TCTGAGGGT CACAGCCAG ATCCCCCAGG GGGTGTATG	780
5	GGCAGGCGCT GTACCTGCAT CCCTTAGTTC GGCAGTGTG GAGTCAGTGC TCCGCGATGT	840
	TGGACTCAAT GAATGTCACA AGGCTGTGG GTCCTGCTG GAAACTCTAG GGCTTCACC	900
10	CACCTGCTG CACCTGCAGA GGGGAATCTA CCGTGAATA TTATTCCTGA CAATGCTGC	960
	TCTGGCAAG GACCACCTGG ACATAGTGG CTTCGATAAG AAGTACAAT CTGCTTTAA	1020
	CAAGCTGGC AGGAGCATGG GGAAGGAGGA GCTGAGGCAC CGGCGGGCG AGATGCCAC	1080
15	TCCCAAGGCC ATTGACTGCC GAAAATGTT TGGAGCACCT CCAGAATGCT AGAGACCTTA	1140
	AGCTTCCTTC TCCAGCCTAG GTCGGGAAG TGAGGAAGAA GGGATCTAG ATTAAACTG	1200
20	CTTCCCTGTT GCTTCATG ASTTGGGAC AGGCTGGGA GGATGCCAG TCAAGGCTC	1260
	CAAGCGAGA CAACAGGAG AGGATCCAC TTTACCAA AGTCTGATT CCCCATCAG	1320
	CAACCTACCC AGTTGTTCG TCGTATGTT GGGGAGATC TGGGCGAGT TGTACAGCT	1380
25	CTGTTCTTC CTGTCTAT AGCGGAAC CCCCTCCAG GTACCCACAG ATCTGCATTG	1440
	CCCTGGTCAT TTTAGAAGT TTTGTTTAA AAAACAACG GAAAGATGA GAGCTACTGA	1500
30	GCCTTTXCC TGAATGGAG GTAGGATGT CATCTCCAC CAATAATGT CCCTCTGCC	1560
	TGACGTTGCT GAAGGAGCC AAGCTCTCC ATGCCCTCT ACCTAAGTGT TGTATTTTA	1620
	TTTTAACTTA TTTATTTGG AGCCACAGCC CCGTGTCTA TGAGGTTCT ATGGAGAGTG	1680
35	AGAAAGGAA GGGAAATAG GCACCATGT CCGGTGTTT GACTTCCTT CAAAGTCAGG	1740
	CACCTGAGC TAGAGGAGT TCAAGCTCC CTTAGGAAGA ACTGCTGCC CCTTCAGTCC	1800
40	TAATTTTTCT TGCTGCCCC GCCTTGGGA ATGCCCTACC CACCTAGGTC CTGACCTGTG	1860
	CAATAAGGAT TGTTCCTCC GAATTTTGT TGGATGTAA TATASTAAA GCTGTTCTG	1920
45	TCTTTTAAA AAAAAAAAA AAAAAAACT	1950

50 (2) INFORMATION FOR SEQ ID NO: 132:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 990 base pairs
 (B) TYPE: nucleic acid

60

TTAAAGATT AAAATAGTT TATATTTT TTTATATG AATATATA TTTAACTAA

384

CTGAGTGGCT TATTATTATG AATTTTCTCT TATTATTTCT ACCAATGCTT CTTATATTAA 120
 AGCCTGATCT TTTCGATATT AATATATGTA CATTAGCTGC CTGTGGATTA ACATTTCCAT 130
 5 GAAAGTATTT TTGGCATGCT TCGATCTTAA ACTTTTTGTG TCTTTATATA AGGTATGCTY 240
 GTTTTACGCA TGATATTTCT AAGCACAATA GTTGAAAGAC AATCTYCACC TTTTACTTGT 300
 ATATTTCAGT GTTATGTAAT TCTTGATGCA TATTACGTCT TATTATTTAA CCAACCTATT 360
 10 TTTTATTATC TAGGGGATTT TTGAGAAAGC CTTATTTTCT TGTATTAAAC AATATTTTTT 420
 ATCATTGTAT TTCTCTCTAT TACTTAGKAA TAGGKTACYC YAAATATATA TTGTGGSTAT 480
 15 TTTCAGAATT GCAATATGCC TCTTTAATTT ATTAGAGGCT AACCTAAATT APTACTTTTA 540
 CGACTTACTT GAATATTTCT GAACTTTTGA ACATTTATTG TTTTATGCAT TTTAATTCTA 600
 CTTGTATTTT TACTACTTCT AATCATTATT ATTGTMTTAG ACAAGGCCAA ATATATNTTG 660
 20 TTAATATCTT ATTCAGCATT TCTTTCTGTA TTTTATGCC ACTATSTATG CTCAATTTCC 720
 TTCTATGCGA TGAACCTAAT TCAGTACTTT TGTMTTTTAA TCTGTGCAGG TAGCCTGGCC 780
 25 ATTAATTTT TATTTTGGT TTGCTGAAAA AATTGTGTTT ATTTCTATAT GCATACTTAT 840
 GCATATAGAA TCTAGGTTG AATATTTTTT AGTATTTATA AATGTAAAGT CATTWATTKG 900
 30 GCTCTATGCA TTCTGCTGCA GAATCAATT CTCAGCCCAA TAGTTTTTCA TTTTAAATTA 960
 CAGAATTTT TCATGCTCTT GCTTTTAGGA 990

35

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1720 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

45

GTCTGATAAG CGACTGTGGT TATTCCTCTA AAGTTTACTT CAGCACTAAC ACTAGTGCTT 60
 CCGCTGGAGT TTGCAATTTT CCAGCTTTAT ACAGGATTTT CTTTGACTG CAAGAGTCAA 120
 50 GGATATAGAG ACTCAACAGT GACATTTATT GTACAACATC AAGGGGAATA GGATACTCAT 180
 CCACTGGGGA TTATTTCTAT CAAAACATGG TCTTCTTTGA ATAAGAAAAA TACATAGTTG 240
 GTTATTATGG ACTTAAACT GTGTAAATG GATATTCTGA TAAAATATTT GCTGCTCTGT 300
 55 AGASTGTGGA AATCTGAGA ATATTAGCTT TACTCATCTT GAGCTTTGAG GATGTTCTCT 360
 GTAGCCCGAT GGTTCATAT TAACTAAAAA AGCTGGGTAT TGTAATATCT CATTTATAAA 420
 60 AACTCAGATG AGAAGAAAT TTTCTTTGAT GGTGAGACTG TTGTCTTAGT TCAGGAAATT 480

385

5
 10
 15
 20
 25
 30
 35
 40
 45

ATTTAATAAT CCTTGTGTTAC CTBTGAATGA AGGAACCTTTG TAAATCTGAT TTATCGTAAA 540
 ACATGAGCCT TTCCAGAGTC AGCTTAGACA CTBTGTCCG AAATAGCCAT CCTTTGCCCTT 600
 ATGCCAAGGA GGGCCAGAGG GAGGGCCTAG TCTTCCTCTG TTGCTGTACA TATATTGAAA 660
 TGCTTTTTTT TTTTATTTTG CATTTGTTAT CTATAATGAG CTTTCTGAGC CCGTATATTA 720
 TGTGAGACAA ACAGGAGTTA TTATGTTAT AACTCCCTT CCATTCAGSA TTTCTGCTT 780
 GGAGGGAAAT ATGTTGACCT TAGAGAATTG TGAATATTGT TGCAATTCCT GAATATATTA 840
 CCATGIGAAT AATAGAGACT GTTGTGCTCT CTAGTATAAG CTATATTTAT TTTGATTC 900
 TTTGAATTAC TAGTTATAAC TGGAGAAATT TTGTTACCTC TATCCTGGCT TGCTGACTG 960
 GCTGTATAAT ACCAGCAGCC TCTTTTAGAG CATCTTAATG AAAACATGGA TGAAGGAAT 1020
 TAATGATGAT ATCTGCAGAC TCCGTAGAAA ATGGCTTTTG TTCCAGCCT TAACATTTTC 1080
 TTCTCAATCA CATTTCAATG TTTGTGAGA GTGGCAGATT CACACCAGAA AACTAGGTG 1140
 TTCATATCCA TAGCATGGAT GCAGAATAAG CAGTTGGGAG AGAAGCTTCT TCCTACCTGG 1200
 TACTCCTCCC ATTCACCTCA GCCCAGCCCC AGACAGGCGT TAGCATTGAG TGTGGGCCCT 1260
 CAGGCAGCCC TGAAGCCTGG CTGGGTCATC AGATGGGGGC AGCCTGTGAC GGGCACCAGC 1320
 GGCTCATTC CAGGAAGAG TTCTGGAGG GTGTTGGCTG TTTTGTAG CTCAGTTTTT 1380
 TTCTGGGCTC CACCATTCCT AACTCCAGGT ASACAAGATA GATGTCACAC ACAACAATTT 1440
 TAAAGTATTT TGCTTAGTCC ATTTTGTTTA TGATTGCAAT GTTTGTTTCT TATTTAATAG 1500
 GCTTTTACT TCATTCATTT AAATTTTAGT GTTTAGAASA GCGGGTACT GTCACGTGT 1560
 AAAATATGTA ATATTTTATA TGTATACCA TGTATATAT ACTTGCAATA TCAGACCTTG 1620
 CATTCATAT ACAATGCAAT TGAATCTTG CAGACCTGCA TTTTTCATG AACATAAAA 1680
 AGATTGTCTG GCACTCCAAA AAAAAAAAAA AAAAAAAAAA 1720

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:

- 50
- (A) LENGTH: 705 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

60

TTTTATTAAT AATCTGAAAT TAAATCTGAT TTATCGTAAA TGTGAGACAA ACAGGAGTTA

1720

GTGGAATCAC AGACACTCCT AGAGGAGAAT GCTGTTCAAG GAACAGAAAG TACTCTTGGG 180
 TTAAATATAG CACCTTTTAT TAACCASTTT CAGGTACTTA TACGTGTATT TTGAGACCTA 240
 5 TCTTCATTGC CCTGTATACC TTAAAGCAAG CCAAGTGBAAC TCTTAAGACT AGATTTAATG 300
 ACTCCGTATT TGAACACCTC TAACAGAGAA GTAAAGSTAT ACGTTTGTNA AATCTGGGAA 360
 GACTTGACTG CTATTCCATT TTGGGTATCA TATGTACCTT GATGAAGANG ATTAGTTTGG 420
 10 GATACTTCAA GTGAAGCCTC CCACTGGAAA CAAGCTGCAG TTGTTTLAGA TAATCCCATC 480
 CAGGTTGAAA TGGGAGAGGA ACTTGACTC AGCATTGAGC ATCAGAAAAG CAATGTGAGC 540
 15 ATCAGASTAA AGCAATGAAG ASCAGTTTTC CAATGAAAAC TGTGTAAATA GAGCATCAAC 600
 AAGTACAAAA TTCTTGTCTT AATTAGTGGG GGTATATAAA AATTCCTTGT AATGCTCAAA 660
 20 TATTTTTTAA AATTGACATT AATAAAGCAT ATTATAAAG TTTCT 705

25 (2) INFORMATION FOR SEQ ID NO: 135:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 323 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

AGCACACACC TCCTTTAGTT GCTCCTAAGG TCATGTTCAA CATTCGTGGA GTGCATTTTC 60
 35 TCCTCAGGGA CTTTTCAGAG ACCCGGAATG TTTGGTGCCTC ACAGACYCTG GCAAGGATCG 120
 GTATTGCTGT TCCTCASTTT TGCTTGGGGA AATGGAGGST CAGTGACGTT CAGTGACGTG 180
 40 CCCAGAGTCA TGCCATTGGC GGGTGCCCCA GKGMTCCAGG TCTCCAGCAC CCCTCGGCCC 240
 CCTCCTCACC AGGTCACATC ATCTCCTGGA TTAGAATCTG CTCACATAGT CTGTCTTGAA 300
 45 ACGAAAAAAA AAAAAAAAAA AAC 323

50 (2) INFORMATION FOR SEQ ID NO: 136:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 582 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

GGACGGAATG GTGCAACCCCT CCTWAMTTTT CTGKGCTGT TGACAACAGA GGGAGGAGG 60

387

GAAAACATTT TTYGTGGGAG AATCCTACTT CTGCAGSGGA GGCCTTAAGC GATGATTTT 120
 GAATCTEGAG CCTTTACCAA CTAATTTTGA AGGAAGATAC CTGGGAAATA TTTGGCATTG 180
 5 AGTGGGTTAT TGAAACAGCA TTAATGAATT CATCTAGAGA ACTCTTTTAT TTATTCAGGC 240
 AACAACTGTA CAAATTGGAA ACCTTSTTAC AGTCCAGTTG TGATTTTGGG AARGTATCAA 300
 CTCTACACTG CAAAGCAGAC AATATTAGGC AGCASTGTGT ACTATTTTTC CATTATGTTA 360
 10 AAGTTTTCAT CTTCAGGTAT CTBAAAGTAC AGAATGCTGA GAGTCATSTT CCTGTCCATC 420
 CTTATGAGGC TTGGGAGGCT CAGCTTCCCT CAGTGTGAT TGATGAGGTT CATGSAITAC 480
 15 TCTTGATAT TGGACACCTA TCTGAACCTC CCAGTGTAA TATAGGAGCA TTGTAAATC 540
 AAAACCAGAT TAAGGTTTGA CTGCTTTCAT TTGATTTTGA AG 582

20

(2) INFORMATION FOR SEQ ID NO: 137:

25 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

TTGGGCAGAG CCCTTGGCGG CTCTGAATA CCTGCHTCT GTAGCGCTAG TTCTCTTCAA 60
 GATTTGCTTA GTGTCAATTC ATTTCGGGTT CTTTCTCGC CATGTTTTC TGTGGGAATT 120
 35 ACGGTCGTT TTGGTCTAT GTACTCTCTA AAATGTTATC GTTTTTCATT TGTCTACTAA 180
 TTTCTGTGTA TTGTTACTA CTGAGTTTCT TAATATCTGA CTGGCCCTCC CCCACGGGCT 240
 40 CTGCAGANDA TAAAATACTC AGGCTGATGG TAGTGCAGAG ACTCTCCCTC CTTGATCAGC 300
 CCAAAGCTTG GTTTCAGGCT TGAGGGATGG AATACATTT TCTTGGTTGT GTGAAGCGGG 360
 CTGTCGATTC TGTAGAGGTG GGTCTAGAGC CTTAGGCTCC AGCTATTTTG AGTTCAGAAC 420
 45 ATCTGAGGCT GTGCTCTCTT CTTTGTATC CAGTATAGG AGAGTACTCA CTGACAGCT 480
 GTGATTTGGG ACTGTTTTTC AGCCCTTGTG GGGGGTGGC CGGAGTCTAC TGGCAAAAAG 540
 50 GACTCTCTCC TGSAGTCCAG AGCACCTTGG AACCAAGTAC AGCGAAGCCC ACTSAGTTCA 600
 GTTGGGCGGG GACACAGAAG CAGCAAGAG CAGGCTAGA AKAGTGGGG CAGGAGAGG 660

60 TCAATTTTCT CAGCTCTCTC CAGTATAGG CAGGCTCTC AGGATTTT CAGGTAAGG 720

	AGCCTTAGAT AGCAGCAGAA GGCTTTTGG ATTCTCTCC TTGAAAAGAT TCTCAGTTAC	960
5	CAAACGTCTC CACCTAGAAA ATAAAAATAC ATTAAGATGT TGANAAAAAA AAANAAAAAA	1020
	A	1021
10	(2) INFORMATION FOR SEQ ID NO: 138:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 1777 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
20	(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 138:	
	CCGAAGATGA TGGCTTCAAC AGATCCATTC ATGAAGTGAT ACTAAAAAAT ATTACTTGGT	60
	ATTCAGAAAG AGTTTAACT GAAATCTCCT TGGGGAGTCT CCGATCCTG GTGGTAATAA	120
25	GAACCATTC AATACACATG ACTAGGACAC GAGACAAGTA CTTTCACACA AATTGTTTGG	180
	CAGCTTTAGC AATATGTCT GCACAGTTTC GTTCTCTCCA TCAATATGCT GCCCAGAGGA	240
30	TCATCAGTTT ATTCTCTTTC CTGTCTAAAA AACACAACAA AGTTCTGGAA CAAGCCACAC	300
	AGTCTTTGAG AGTTCTGCTG AGTTCTAATG ATGTTCTCTT ACCAGATTAT GCACAAGACC	360
	TAAATCTCAT TGAAGAAGTG ATTCTGAATG TGTTAGAGAT CATCAACTCC TGCTTGACAA	420
35	ATTCTCTTCA CCACAACCCA AACTTGATAT ACGCCCTGCT TTACAAACCC GATCTCTTTG	480
	AACAATTTCT AACTCATCCT TCATTTCAGG ATATAATGCA AAATATTGAT CTGCTGATCT	540
40	CCTTCTTTAG CTCAAGGTTG CTGCAAGCTG GGAGCTGAGC TGTCAGTGA ACGGGTCTCTG	600
	GAAATCATT AAGCAAGGCT CGTTGCGCTG CCCAAAGACA GACTGAAGAA ATTTCCAGAA	660
	TTGAAATTC AATATGTGGA AGAGGAGCAG CCCGAGGAGT TTTTATCCC CTATCTCTGG	720
45	TCTCTTGTCT ACAATCTAGC AGTGGGCTG TACTGGAATC CACAGGACAT CAGCTGTTC	780
	ACCATGGATT CCACTGAGG GCAGGATGCT CTCCCACCCT GACCTCTCCA GCCAAGCAGC	840
50	CCTTCAAGTT CTTTATTTTC TGGTAACAG AAGTAGACAG ACAGTTTACT TGTGTATCT	900
	TCTGTTAAAG AGGATGTCAC GAGTGTGTTT TCCTCACACA CTTTGAATTG GAGAATTGGT	960
	GTTAGTTGGC AATAGATAAC TCAGGCTAGA TAGTATTGCA AAAAGGGGAG GAAATACACA	1020
55	ACAATAATAA ATGTAAAAAC CTGCTATTCA ACATGCAGTT TTATTTTCGAR GCCAAAAATC	1080
	TAGAGCTTTC CCAAGATCCT GTTGCCTTAG GCACATNCAC ACTTCAACAG TGCACTAT	1140
60	CCAACAGTGC ACACTATTCA ACAGTGACA CTATTCAAAA CCGTAGACTA TTTTATTGCA	1200

389

TGTTCAGAT ATTTGTTTTG GTTTTATGTG TGTGTGAGAG AGAGAGATTG CTTTGACATT 1260
 AAGGAGTATC AATGAGAAAA GATGATCAGG CAGGAATTAA TAAAGAAATG AAGTCGTGTG 1320
 5 TGTTTGTTTG CCTGTGAGAG GGCACACAAT TTCATAAACA CCATGCCTGG ACAATTTGAT 1380
 ATTAATATTT AACACCTCTG CATCTTTTTC TTAAAAAAGA ATATGGGCCA GATACAGTGG 1440
 CTCACATTTG TAATCCGAGC ACTTTGGGGA GCCAAGTTAG CAGAATCCCT TGAGCACAGG 1500
 10 AATCTGAAAC CAGCTTGGGC AACATAGTGA GATCCCATCT NTACAAAAAA CTAAAAAAT 1560
 AGCCAGGAT GATGACACAT TCCTGTASTC CTAGCTACTC AGGAGGCTAA GSTAGGAGGA 1620
 15 TTGCCTGAGC CAGGAGTTC AAGGCTGAG TGAGTAAGN ACGTGCCAGT AACTCCAGC 1680
 CTGAGCCACA AACTGAGACC CTCTCTGCA AAAAAAAN TTAAAAAGTC GGGGGGGG 1740
 CCGGTACCA AATGCGCGGA TATGATGTA AACAATC 1777
 20

25 (2) INFORMATION FOR SEQ ID NO: 139:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 643 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

35 TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTGGG AATGAGAAAA TAACTTTATT 60
 TTCATTGTGG GGAGCGGGCC GATCTCCAGC CTCAGAACTT CTGGAAGTGC TTCTTGGTGC 120
 CGGCAGTCTT GGTGACCTTG AGCAGCTTGA AGCGCACTGT CTCTCTCAGA GCGCGGCACT 180
 40 GCGCCACTGT GACATGTCA CCGATCTGSA CGTCCCTGAA GCAGGGGGAC AGGTGTACAG 240
 ACATGTTCTT GTGCGCTTC TCGAAGCGGT TGTACTTGGG GATGTAGTCC AGATAGTCTC 300
 GCGGATGAC AATGCTCTC TGCATTTTCA TTTTGGTCA CCACGGCAGA GAGGATCCGC 360
 45 CCTCGAATGG ACACATTACC AGTGAAGGGG CATTCTTCT CAATGTAGGT GCGGCTCAAT 420
 AGCCTCCTTG GCGTCTCTTT GAAGCCGAGA CCGATGTTCT TGTAGTAAC CCGCGGGAGC 480
 50 TTCTCCTTGC CAGTTTCTCC CAGCAGGACC CTCTTTTCTT TTTGAAAGAT GGTGCGCTGC 540
 TTTTGGTAGG CAGGCTCACT CTGAATGTC GGCATCTCT CCGTGGGMAV TCCTGCAGGC 600

60 (2) INFORMATION FOR SEQ ID NO: 14

390

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1220 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

GGCACCAGGA TGATAGACCT ACTGAGGAA TACATGGTTT ACAGGAAGCA TACCTACATF 60
 10 AGGCTTGATG GCTCATCCAA GATCTCGGAG AGGCGAGACA TGGTTGCTGA TTTCAGAAC 120
 AGGAATGACA TCTTGTGTT CCGTTAAGC ACAGGAGCTG GAGGACTGGG TATCAATCTC 180
 15 ACTGCTGAG ACACAGTGCA TTTTCTATGA TAGGACTGG AACCCCACTG TGAACAGCA 240
 GGCCATGGAC AGGGCCDACC GCTTAGGGCA GACAAAGCAG GTTACTGTGT ACCGCTCAT 300
 CTGTAAAGGC ACCATTGAAG AACGCATTCT GCAAAGAGCT AAGSAGAAGA GTGAGATTCA 360
 20 GCGGATGCTG ATTTCAGCTG GGAAGTTCAA ACCAGATACC TTGAAACCCA AAGAGGTGGT 420
 TAGTCTTCTT CTAGAGGAGG AAGAGTTGGA GAAGAAACCT ATGTACTCTA AACCTCTATA 480
 25 CACTCCCCCTC AGGTATCTGA GAATGGAAGA GGTACTTGGG TGTGTGCCAA GGSTTAGGCA 540
 AAGCCAGAGG CTGTATTTAG GGAAGTATT TTTGTGCTCA TATTTTATAT AAAAACCCAA 600
 ACAAGAATGT GTTTGTAGGC CAGGCGTGGT GGGCTCGGCC TCTAGTCTCA GCATTTGGGG 660
 30 AFGCCAAAGT GGGCAGATCA CCTGARGTCA GGAFTTTGAG TTTGARACCA GCCTGCCCMA 720
 CGTTGTGAAA CCCCACCTCT ACTARGAFTA CCGAAAATTG GTTGGGCATG GTGGCGGGCA 780
 35 CCGTAATTG CAGCACTTTG GGAGGCTGGG GCAGAANAAT TGCTTGAGCC CAGGAGGTGG 840
 AGATTGCGST GAGCCGAGAT YGTGCCATTG CMTCCAGCC GGGCAATAA GAGTGAAAYT 900
 CCATCTTTTA AAAACAAACA AAAACAAAAA ACACAAGAGC GCTCACACCT GTATCCCGAG 960
 40 CACTTTGGGA RGCCGARGCA GGTGATCAC GAGTCAAGG GTTCCAAGAC TAGCCTGGCC 1020
 AACCTCGTGA AGCCCCGTCT CTAATAAAAA TACMAATATT AGTGGGGCGT GGTGGTGGGC 1080
 45 ACGTGAATC CCAGCTACTC GGGAGGCTGA GGCAGGAGAA TCCTTGAAG CTAGGAGGCA 1140
 GAGGTTCAG TGAGCCAGGA TCGTGCCATT GCACTCCAGC CTGGACAACA AGAGCAAGAT 1200
 50 TCCATCTCAA AAAAAA

55

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 AATTGGGCAC GAGCCAGGTT AGCTGGAAGG GCAGCTCTCC AGGCCCTGCC CACCCACAG 60
 GGGGCTCCTT ATGCACAGCG GGGTGTCTCC TTGTGGCCAT AGAAACGGAA CTGGCTCTTT 120
 TCAACASTGC TGCAAGAGGA TGTTTATTTA ACGTGGCCC CCAAGGAGGA AAGGCACAGA 180
 10 CYTTCCTCCC TCCTGGAACA TCCAAGGCCA CTGATCTCTC TGTGTCCCTC TGATATGGGG 240
 TGCCACTCCA CCAAGAGCAC CACGTGTGCA GCTGAGTCCC AGAAGCTTGA AGAAGAGYGC 300
 GAGGGAAGAG AGCCAGCTCT GGATCCGGC ACCCAGGCAG CAGACTGCAA GGATGCCCCG 360
 15 CTGAAGGATG GAACCCCTGA GCCAAAGAGC TGAATGCCCT CTCTCCAGAG TCGGACCCTC 420
 ACCTCYTTCC TGGAAGTCC TTGTGCCCTA GAACCATGAG ACAATCCCCA CCTTGAGAAG 480
 20 CTCCGATCAC TCGGAGGAGA GAGAAAGGCT CCAGCTTTTG GATTCAGGCT TCAGAAGTTT 540
 TTAGCAGCCT TTCTCATIG GAGAGGTGGG GAAAGGATAA ATTCTTATA AGGAAATCCC 600
 TAATTTCCCC CAGCTCTCTC CCHCINGAAG AAGGAACNAA AGAAAGTTCC TTCCACACGT 660
 25 TTTGTGGAA ACTTTTCTCT TGCCAACCTT CCTTGGATTG CCAGAACAAA GGCCTCCAGA 720
 A 721
 30

(2) INFORMATION FOR SEQ ID NO: 142:

- 35 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1463 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

45 ATGAATTAAT GTTTATAAAT GATGTACTG AATTAAAC CCTACAGTTT CATTGCAAT 60
 TTGACATTAC TTTATTATAC AATTGCAAT TAAAAGGCTG CACCAGTTG GTTTTCTCT 120
 GTTTTATCT CAAATATAG AGATTCTGT ATTTATTTC CTTTITATG TATTAAAAAG 180
 50 AAAATTCTAA TATAAGAT TTCAATAGGA TGCATAGGA TATTACGTTT TTAAATGCT 240
 TTAGATCTGT GATTCTGAC TTAATATTTA TTTTATCCC TTAAAGTCAG GGATGCTTTA 300
 TTCTATTTTA AAGCACTTAT GAGTTACATG TTGTAATCAA GTTGGACAA TATATTTATC 360
 60
 ATACAGTTAT TAAATCTCT TACATTTAAA ATTCTTTT TATATATAG AGAATTAAG 540

	GAAAAAAAAA TTCTCATTAT TTGCAAAGAA TGAACAAGTT AATGAACAAA CAAACTAGAT	600
	TTGGTATGTT TDCAGCTTTT GTATCATGTT TAATTGTTTA ATTTGGTTGA AAAACTGCAG	660
5	TTGAGAAATC AGATAGCAAT ATAGACATTC ACAGCAGCTC TGTGGATACC ATGTAATTGT	720
	CAGSTAATTT CAGAATGTTG AAAATTATTC AGTGCAGCCG TCATAGTATC ATACTTGAAG	780
10	AAATTGATTA CAGTTCACCT AAATTGTTGA AGATAAATTA TTTTAAAGG TTATGAAAAC	840
	TAAATTATAT TAATTCATAT GTTTGATTTT TAAATCCCAC CTCCTCAAGC TATCCAAATT	900
	NCTGACTTTG AAAATAACCA TGAGAGATGC CACATTTCTC TCTGGGAAAC TACCACTCAA	960
15	AGAATAATTC TTAAAAATTA AGCTTTTAGG TATTAGAAGC TETTATAAAG TATAAAATTA	1020
	AGATATAAGC AGATCACATG TAAATCATTC CTAAAGCACA AGAAAAAAT GTGCCTTGAT	1080
20	GTACATATAT TACTAAGTTG CTTCTCCCAG TTTACTTTAA AAATGGCTTT AAGGATAAAG	1140
	AATAAATGTC ATAGCTGTGC ATGCATTATA TATTTGCATT TSCAAATTTG CCATTGTTTT	1200
	AACAGCTGTC TGGCTGACTT TCAATTTTAA GACGTGAATT GACATACAGC CCATAACTTT	1260
25	ATAATGGGTC CTCATTTATC TTATCTTTCA GTTAGTGGAA AAACATTTCA ACCTGACTAA	1320
	AATTTGGAAT TGTCTCTTTT ATGTTCCATC CTCTGTTSTT ACTAGATTTA GTTTAAAAAT	1380
30	TGTGTATGAC CATTAAATGA TGTCAAAAC ATGTAAATAA AAGATGTGA ATCTTGTTGA	1440
	AAAGCAWRRA AAAAAAAAAA AAACCTCGA	1468

35

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 300 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

45

	TGAATTTTTT GCCAAACTTA STAACCTCTGT TAAATATTTG GAGGATTTAA AGAACATCCC	60
	AGTTTGAATT CATTTCAAAC TTTTAAATT TTTTGTACT ATGTTTGGTT TTATTTTCCT	120
50	TCTGTTAATC TTTTGTATTC ECTTATGCTC TCGTACATTG AGTACTTTTA TTCCAAAAC	180
	AGTGGGTTTT CTCTACTGGA AATTTTCAAT AAACCTGTCA TTATTGCTTA CTTTGATTAA	240
55	AAAAAAAAAA AAAAAAAAAA AAACCCCNAG GGGGGGCCCG GGTNCCCAAT CCCCCCAAA	300

60

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2243 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

10 TCCCTCCCTT CCTCAGATT GTGACAGTA GTTCTCAGC CTGCACCTG GATTCCTTCT 60
 TCCCTTCTT AGCTCCATG GACTCGCCG AAGACTGTG CTCACAGAG CACCACCCCC 120
 TTAATCTCA AGCTCTGACT GTGAGTTG TAGATGCTT TGATCTCAG TATTCTTCT 180
 15 GCAATGTTT CAGCTCTTCT CCTTCTGGG AGCTGGCTG ATAACCTGAT TTCCCCAAA 240
 CTTTGTGAA TCTCTCTGCT CTTTACCCA CCGAGGCTCT TGTGTGGTA TGACTGTAGA 300
 GATGGGGGT ATTCTAGGG TGGCCCTCC CAGGCAGGC CCGTGGAGCC TGATGCTACT 360
 20 CTTATCCACT GCAATGTAAG GTCCCATGC CCAATCTCT GCACTGTGC ATGTGACCG 420
 CCGAGTCCC TTYGGCCCT CTCAGCCCT GTCCTGACT GAGCTGACC AGCTACTGTT 480
 25 ATGGCCCTT TCCCTCTG TAGGCTGGCA AGCATGGCC CAGGGGCCG CACCTGGCG 540
 CCAGCTGCT CCGTTCGAC TATCAGCCCT GTCCTATGG CTAACAACA ACCTGGTGT 600
 CTATCTTCA GCTTACATG ACCCCAGCA CTACCAGGT CTGATTAAT TCAAGATTG 660
 30 AAGCACAGT GTCTCTACT GCTCTGCTT CCGCCACCG CTCTCTGTC GTGAGGGGT 720
 AGGCTGCTG CTCTCTATG CTCCGGAGC CTGCTATGA CAGCGGGCC TTCAAGTTCC 780
 35 CCGGAACAC CTTCTCAGT CCGCTCCAG ACCTGCTGC AGCCCATGC CCTGCAAT 840
 CACTCCGCTA GGGTCTCTG TCTCTATTCT GTACTGCCCT ATCTCAGGCT TGTCTCAGT 900
 GTACACAGG CTCTCATGA AGGACAGNG GCTGCCCTG GCACTTCAG ACCTCTTCT 960
 40 CTACACTTTT GGTGTGCTT TGAATCTAG TCTGCATCT GCGGGGGCT CTGGCCACG 1020
 GCTCTTGAA GGTCTCTAG GATGGGAGC ACTGTGCTG CTGAGCCAG CACTAAATG 1080
 45 ACTGCTATG TCTCTCTCA TGAAGCATG CAGAGCATC AACGCTCTT TGTGCTGCT 1140
 CTGCTGCTG GTCTCAAGC CCGTCTCTG AGCACTCTG CTAGGCTCT AGCTCAGCT 1200
 CCGCTCTTT CTGCTCAGT TCTCTATTG CTTGCCATG CGCTGTACT ATGCTAGCC 1260
 50 CTAGTCTCT ACAACTTCA CCTGATTCG GACCCCTGA GATGCGGCG CACCACAGA 1320
 TCCCTCTCC AGGCTTCTT CCTCTTCCA TCACAGGCC TATAACAAT GCTTGTGAG 1380

 60 ACAAATCTT TCAATTA ACAAATTA TAACTAAT AATTAAT TCAATTAAT 1440

5
 10
 15
 20
 25

CCCATCCTTG TGGGGCAGCT CCTGCTTTG TCCTGCATGA ACAGAGTTGA TGAAGTGGG 1620
 GTGTGGGCAA CAAGTGGGTT TCGTTGCTTA CTTTAGTCAC CCAGCAGAGC CACTGGAGGT 1680
 GGCTAGTCCA GGCAGCCAT GGTGCATGAC TCTTCCATAA GGGATCCTCA CCTTCCACT 1740
 TDCATGCAAG AAGGCCAGT TGCACAGAT TATACAACCA TACCCAAAC CACTGTGATA 1800
 GTTCTCTCCA GTTCAGCAA TGGCTAGAGA CATGCTCCCT GCGCTCTCA CAGTGTGCT 1860
 CCGCACACCT AGCTTTTGT CTGGAAACCC CAGAGAGGCC TGGGCTTGAC TCATCTCAGG 1920
 GAATGTAGCC CTTGGGCGCT GGTTAAGCC GACACTCTG AGTCTCTGT TACCGCTAG 1980
 GGCTGTCTTG AAGCCCGCTA CCGACTCTGA GGCTCCTAGG AGGTACCATG CTCCCACTC 2040
 TGGGGCTTGC CCTGCTTAG CAGTCTCTCA GCTCCCAACA GCGTGGGAA GCTCTGCACA 2100
 GAGTGACCTG AGCCAGGTA CAGGAACCT GTAGCTCAAT CAGTGTCTCT WTAAGTGCAT 2160
 AAGCAATAAG ATCTTAATAA AGTCTTCTAG GCTGTAGGCT GGTTCCTACA AACACAGCCA 2220
 AAAAAAAAAA AAAAAAATC GAG 2240

(2) INFORMATION FOR SEQ ID NO: 145:

30
 35

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1082 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

40
 45
 50
 55
 60

GCCAAGCTCT AATAGACTC ACTATAGGGA AAGCTGTAC GCCTGCAGKT ACCGGTTCCG 60
 GGAATTCCCG GGTGACCCA CGCGTCCGCT TCCGTGTGTC AAAATCCTCA CCTCCTTCAT 120
 AACCATCTCC CACAATTAAT TCTTGACTAT ATAAATTAT GGTMTGATA TATTATCAAT 180
 TTGTAATCAA TTGAGATTTC TTTAGTGTCT GCTTTTCTGT GACTCAACTG CCCAGACACC 240
 TCATGTACT TGAAAAGTGG AACANCTTGG GAATGCCATG GGSTTTGATA ATCTGCCAGG 300
 GACATGAAGA GCCTCAGCTT CTTGGGACCA TGACTTTGGG TCAGCTGATC CTGNACATGG 360
 GAGAAACAAC ACATTTTCT TGTGTGTGC TTCTAGCAGC TGTGGGGAG GACCKTGACC 420
 CAAYAGTGT CCCATGCTGT TTCTGTGAA ATGCTCTCGG CTATGTAGCA GCTTTTGATT 480
 CCTGCATAC CCTAGGCTGC TGCCCTATC CTGTCCCTTG TTTATAACAT TGAGAGGTTT 540
 TCTAGGCAC ATACTGACTG AGAGCAGTGT TGAGAAGTGG GGGAAATGG TGAAGTCTT 600
 TAGAGCAAGG CTGGGCATCA GCACCTGTCC AGCTCTACTT GTGTGATGTT TCAGGAAGTC 660
 AGCCCTTTT TCTGCCTAGG ATAAGGAGCT GAAAGATTAA CTGGATCTY CTAATGCTCC 720

AAATCTTTTG GTCACAATAA AGAGTCTCCA AATTAGAGAC TGCATGTTAG TTCTGGATGG 780
 ATTTGGTGGC CTGACATGAT AGCGTGGCAG CTGTGAGGGG ACCCGGTTT TAAGATGCAT 840
 5 GGGCAAGGTC TGTGCAATG GAAATGCTTA CACTGGGTGT TGGAGATGTT TGCTACCTCC 900
 TGCTATTTT GGGTTTTG TTCTTCACT ATGGTAGGAC CCGTGGCCAG CATTGGGCT 960
 10 TGTGATGTCA GGGCCATTGA CTACCTTTC ATGCTCTGAG GTACTACTGG CTCTGCAGCA 1020
 CAAATTTCTA TTTCTGTCAA TAAAGGAGA TGAAATRAA AAANAAAAA AAAAACTGG 1080
 NG 1082
 15

(2) INFORMATION FOR SEQ ID NO: 146:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4313 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

CAAGCTAGTT TGAAACTAGG GGTGGGGTTC GCGCGTCGTC GTTGTTTGTC GCGGCATCCC 60
 30 CGCTTCGGGG TTAGGCGCGT CCGTGGCGCC CCGTCTCTC CTGCTCTGGG ACCCATAGAT 120
 CTCAGGCTCG GCTCCCGGCC GGTGGCAGCC CACTGTTGAC GCGGCGCGTA CTGCGGCCCC 180
 35 GTGGCCAGCA TGTCCCTGCA GGGCAACGG AAGGAGATCT ACAAGTATGA AGCGGCTGG 240
 ACAGTCTAGC CATGAAGTG GAGTGTACGG CCGGATAAGC GTTTGCGTT GCGGCTGCGC 300
 AGCTTCGTGG AGGACTACAA TAACAAGGTT CAGCTTGTG GTTTAGATGA GGAGAGTTCA 360
 40 GAGTTTATTT GCAGAAACAC CTTTGACCAC CCATACCCCA CCACAAAGCT CATGTGGATC 420
 CCGTACACAA AAGGCGTCTA TCCAGACCTA CTGGCAACAA GCGGTGACTA TGTCCGTGTG 480
 45 TGGAGCGCTG GTGAAACAGA GACCAAGCTG GAGTGTTTTC TAAACANTAA TAAGAACTCT 540
 GATTCTGTG CTGCGCTGAC CTGCTTGGAC TGGAAATGAG TGGATGCTTA TTTTTAGCT 600
 ACCTCAAGCA TTGATACGAC ATGCACCATC TGGGGGCTGG AGACAGGGCA GGTGTTAGGG 560
 50 CGAGTGAATC TCGTGTCTG CCATGTGAAG ACCCAGCTGA TCGCCCATGA CAAAGAGGTC 720
 TATGATATTG CATTTAGCG GCGCGGGGGT GGCAGGGACA TGTGTGCTC TGTGGGTGCT 780

	CTGTSGCCAG GTTAAACAAC CATEGAGCAT GTGTCAATGG CATTGCTTGG GGTCCACATT	1020
	CATCCTGCCA CATCTGCACT GCAGCGGATG ACCACCAGGC TCTCATCTGG GACATCCAGC	1080
5	AAATGCCGCG AGCCATGAG GACCCATCC TGGCTACAC AGGTGNAAGC WAGATCAAC	1140
	AATGTGCACT GGCATCAAC TCAGCCCGAA YTTGTGCAAT CTGCTACAAG AACTGCTGG	1200
10	AGATACTCAG AGTGTASTGT TGGTGGGCT GTGCCAAGG GGCAGGGGCT TTTTATTTT	1260
	CTGCTCTGC CCCACCCCA AGTAAGAAG AAACATSTT CCAGTGCCA STATGTCTT	1320
	CATTGCTTTG CACCCACTGT TACCAGAAGC TGCTCTAGG GTTCTGCGC AGTCACCCCA	1380
15	TGCCCTCTG TGGCAGACTC AGTGTGTGT GGGGCTCTT CAGCCAGAG CTGASTTTTA	1440
	AGATTTTCTG TCTTTCTCT TTCTCTTTG GTTCTCAAT TAAAAATGT GTGTATATT	1500
20	GTTTGTGAG CGTTGTGTG AGGAGCAGTT CAGCACTGG CTGTGTCTAT TCTTGTCCC	1560
	AGGTGTCTCT GTTTGCTGCC CAKSYWKKT TTTCTGTCT CTTCCATCTG CATGTTCTGT	1620
	TTAGCACTWA CGTGGGAACA AATACCAAT TGTCTTTCT CCTAGTATCA CTGTCTTTAA	1680
25	CAAAATTTAA CTTTGTATAT TTGTATCTA TCAGGCTAAT TTTTATAA AAAGAATTT	1740
	ACTCTCTCC TTCATTTCTT TGCTTATAG TCTCTCTCT TTGCACCTC TTCTTTCCC	1800
30	TCAGTGCTG GAGCTGGTAC TGGGCCCCG GCCCCATGAG TAGTTTCCCT TCTTCACTCA	1860
	CTGCTCTGT ACTACATACC TGACCCGGAG TCCAAACCAC CTGTGTCTC TGAAGTCCAC	1920
	TGACTCATCA CACCTTTCTT AGCCTGGCTC CTCTCAAGG CATCTCGGC TTGTAAACAG	1980
35	ACATAGGAAG CCTCTGTTTA CCCTGAAGCA CCACTGTCCA GGCATCTGT TCCCACTGGC	2040
	AGCATGGTAG AGCTGAGAGA AACAGGCTCT CAGGCTACCT GACTTGAGG GATCTTTTC	2100
40	ATGAAGCTGA ACTTCAAGCA TATTCCAGT ACATTCTTT AGAGTCTGT TTTCCATCCA	2160
	AATATAAGCC CCAGGCCATT CCACCTAGTG TCTTTTCAAT GATAGGCAAG AATGATATCT	2220
	GAGTTGAACT TCGGTGCTTC TGTGTMTGA GTTACTGTG CCTGGTGGTA TATTGGGCAT	2280
45	TCTTTGGATT GAGTGTCTG AGGTGAGAGA GTCTCCCGA GGCATCCTGT CTGTGCTTCC	2340
	AACCTTGAAC AAGACCTTAC ATGAGAGATG GACTGATGGA CTGCGCAAT CCTGGGCTGT	2400
50	CAAGTGATA GATAGTTAAA AAGCATTATA CTGTGGGTAA TGAAAAGGCA GGAAAAAAA	2460
	AGAAGGAAAA GGAATTATAG ACCCCCAGGG TCAGCCAGTT AAGAGCTATA CCCACACCTG	2520
	TCAACCCCTC TCTCCCCAG TTTAGGTTCT GAGCAGTATT GGACTGTAG CTTGAGTTG	2580
55	TCTTTTGAAT TGCAGGCCGC AGTGTCTTTC TTTATGTGA ATGASTTCA TGGAGGGCA	2640
	TATGTGTGAT TCCACCGTTA GATGAGCCCT TGGGCAAGC AGTTTGGAT GTGTCTTGG	2700
60	GGGAAAGTTG GCTGTTTCTT TGGCTCTGC TCCTACCGA AGTTTTTAAG TCCCTCTGAA	2760

	TTGCTCATCT GAGATTAGTA GAGTAGCAGG CCTAAGGAT GATGOTTING TCCTCTTTGG	2820
	TTCTCAGCTG CTGAGAAGT AAAACAGTAA CTTTGTCTTT CTGGGCCCCT AAGCTTTTTT	2880
5	GGTTAAGTCT TCCTTTTCAG AAGTAGATGT CATTATATGC CAAAAGTCTA GCTCTTTGCT	2940
	TTACCATACA GGGACCTCTC CCAAAGAAAA AGGCTCTTTT TTAGCCAGC ATATTTCGCC	3000
10	TTCTACCCTT TTACTTTGTT GTTCTGATTT TAGGACTCTG GCTGGCCATG TGCTTGTGCT	3060
	TGCTCTCTCT GCATTTCCTA CTGGATTTGC ACTGCATCGT TTGGAGATAC AAACCGAGCA	3120
	GTTCTTGGTC AGAACCTCTC TCTGCTTTTC ATTCTGTTTG ATAATGTTTA CTGGCTCTTT	3180
15	CTCTCAAGGG TAGCAAGGTC AAGCTGATGG CTGCTTGTCT AGGAGGCCAT CAGTTCTTTC	3240
	CTGTGAGAGG GGCTCTCAAA TGGAACTCAG TGGTAGAAGG GGCTGCTCTG CTGGGAGGG	3300
20	CTTACATCCA CTGAGTTCTA AGATTCTTTT CTGATCTGC ACCTACGCTT GGTCTCTATG	3360
	GTGGAAATTT TCAGCTAGAA CTCAGAAACA ACAACTTGAA AAAAAATTA TAATTAGAAG	3420
	ATATTTCAT AAGATAGCTA TTTACTCTGG AAACCAACAA CTTTGTAGAT TTCCCTTGCC	3480
25	CTGTGGAGCG CCAGCTCTCT TCATCTCTTC TTAGGTCTTG CACTAGCTC TTCCCTGAA	3540
	TGCCACCCGG GACCTACGCG GACTCCACCC CCTAAGCAA GACACACAT ACTCATAGTT	3600
30	GATGASTTGC TGGTCTTTGA GTCCAGCTC TCTTACCCTC CTTTACTCTC ACCAGCTCGA	3660
	CGACCCATGA CTGAGTAGGG GATTCTTACA GTCTCAGGAT TTAGAAAGTC TGTAAGCCAT	3720
	CCATGCTCCA GAAAGTAGCG ATCTGTTGTA CTTGCAAAAA CAACTCTGTA AATTGTTGAG	3780
35	GTTCTCAAAC TCACAGCTAG CGAGACTGCG TGGGAGGCCC TGGATCTGTT CTCCCTGACT	3840
	GCGGAGGAG CAGCTACTAG GACTTTAGCA GGAAGCCAC ATGAGGCTC CGCCAGGCTG	3900
40	TGGCCCAGCT GGTGATGCGC CTTTGTCTCC TGGCAGCCTG AGCCACAGCT GCTCTATTG	3960
	TCCTCATCTG TCTGACTGA AGGATGGAGG TCTGAATAA ATTAGGCTC AGGCTCTAC	4020
	CACCAGAGAG CTGGAGAAAT GGTCCACGTC ATTCAAGGAC CTGAATTTT TATGCTTAGG	4080
45	AGCACTGAGG TCTCTCTTCT CCAGGGAGGA ATTAGCTCTC AAGGTTAGGA CTGCAAGAGG	4140
	CAAGSTATTT AATACTTGG CGAGGATGGG TGTGTTGGCT CACACTCTTA ATCTAGCAT	4200
50	TTTGGAGGGT TGAGGTGCTC AGATCCCAAG GTCAGAAGAT CGAGACCATC CTGGTAACA	4260
	TGGTGAAACC CCATCTCTAC TAAAAATACA AAATTAAATT GGGGGGGCT GAA	4313

SEQUENCE CHARACTERISTICS

(A) LENGTH: 1193 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5
GACAGAGGCT CAAGCTGACT TGGATTATGT GGTCCCTTAA ATCTACGGAC ACATGCAGGA 60
GGAGTTGCGG GCGCGGTAG AGAAGACCAA ATCTCAGGAT CCGCTGACTG TGCTGCTTA 120
10 TCAWYGGGG AGTGTCTACT CAGTTGCTAT GGTACAGGC CTCACCGTGT TGCTCTTCC 180
ACTTCTCTGG TTGCAATGCG AGGCAATCAG CCTTGTGTTC CTGCTTCTGT TTCTGCAGAG 240
CTTCTTCTTC CTACATCTTC TTCTGTCTGG GATACCGCTC AGCAGCGCTG GTCTTTTAC 300
15 TGTGCAATGG CAGGCACTGT GGGTTTGGGC CCTCATGGGC ACACAGAGCT TCTACTCCAC 360
AGGCAAGGAG CTTGTCTTTC CAGGCACTCA TTGCAATGCA GCTTCTGTGG GATGCGCAGA 420
20 GGGTCAATGG TCTCTACTTT GGTCTGCTTC TTCTCTAGTG GAGGCAAGCA CTTTGGCTTC 480
CGAGCTCTTC TTCTCAGTAG GTTCTGCACT CCTCTGCTTC TGGCTTTTCC TCTGTGAGAG 540
TCAAGGCTTC CGGAAGAGAC AGCAGCGGCG AGGCAATGAA GCTGATGCGA GATCAGAGC 600
25 CAGAGAGGAA GAGGAGGAG TATGAGAGAT GGGGCTCGGG GATGCGCTTC AAGCTTCTA 660
TGCAGGAGTG CTGAGGCTGG GCTCAAGTA CTTCTTTATC CTTGCTATTC AGATTCTGGC 720
30 CTGTGCTTTC GCAGCTTCCA TCTTGGCAG GCATCTCATG GTCTGAAAG TCTTTGCGCC 780
TAAGTTTATA TTCTAGGCTG TGGCTTTCAT TCTGAGCAGC GTGGGACTTC TCTTGGGCT 840
AGCTTCTGTG ATGAGAGTGG ATGCTCTGTG GAGCTCCTGG TTCAGGAGAG TATTCTGCGC 900
35 CGAGCAGAGG TAGGCTATTC TGTATTACT GGCCTTGGC TACAGAGAGT GCTGGAGAAC 960
AGTGTAGGCT GGGCTGTACA GGTACTGGAT GATCTGCAAG ACAGGCTCAG CCACTCTCTT 1020
40 ACTATCATGC AGCGAGGGGC CGCTGACATC TANGACTTCA TTATTCWATR ATTGAGGACC 1080
ACAGTGGAGT ATGATCCCTA ACTCTGATT TGGATGCATC TGAGGGACAA GGGGKCGGT 1140
45 STCCGAAGTG GAATAAAATA GCGGGCGGTG GTGACTTGCA CCT 1183

(2) INFORMATION FOR SEQ ID NO: 148:

50

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 734 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

60

GAATTCGGCA GAGTGAAGCA TTAGAATGAT TCCAACACTG CTCTTCTGCA CCATGAGACC 60

399

AACCCAGGGC AAGATCCCAT CCCATCAGAT CAGCCTAGCT CCCTCCTGGC TGCTGCCCAC 120
 GATGTCCGCA GCATTACCTT CCACTGCCCT TCTCCCTGGG AAGCAGCACA GCTGAGACTG 180
 5 GGCACCAGGC CACCTCTCTT GGGACCCACA GGAAAGAGTG TGCAGCCAAT TGCTGGGTG 240
 ACCTTTCTAT CTTCTCTAGG CTCAGGTACT GCTCCTCCAT GGCATGGYT GGGCCGTGG 300
 GAGAAGAAGC TCTCATACG CTTCCACTC CTTCTGGTTT ATAGGACTTC ACTCCCTAGC 360
 10 CAACAGGAGA GGAGCCCTCT TGGGTCTTC CCRGGGCAAT AGGTCAAACG ACCTCATCAC 420
 AGTCTTCTT CTTCTTCAAG CTTTCTATCT TGAACACAGC TCTCTGCTT CCTTCTGAT 480
 15 TTCTGAGGCT CACCACTGCC AACTCTAGG AACTAGAGA GCTCTCTGT CTTCTATAG 540
 TTGCTCTGAC TGAGCCTAAA GTTCAGAAAA TGGGTGCCAA GGCAGTCC AGTGTCTTG 600
 GGCCCTTTG GCTCTCCTC ACTCTCTGAG CTTCCAGCTG GTCTCTGAC ATGAGCCAG 660
 20 GACTGTGAGT CTGGGCAGT CCAAGCCCT CACCTTCAAG AAGTGAATA AATGTGAGCT 720
 TTGCTTCTAT TTAA 734

25

(2) INFORMATION FOR SEQ ID NO: 149:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1435 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

GGCACAGTGG ACTCCAGAGT CTTCTCTGCG CTTCTCTGC CTGGGAGAC CCACTGTGTG 60
 40 CATGGCATCA CTGACTCCCA TACCTCTGGC TATCAAAGGT TTCTGCCATG GGCACCTGG 120
 AAGSAAAGCA GAGGAGGTA GATAGGAGA TCAGGTGCTT TATACTCTG TTTCTGCTCT 180
 CTGAAATCT CTGAGGCTGG CTCTCTCCAG AAGGTGCTG GTCTCTCAR GATGCCAAA 240
 45 TCTACAAGAA TCTCTCTCT TCCAGTCTCT ATACCTCTC CTCTCTTTG TCTCTTAGA 300
 CTTTGCATA GTAGCAGCA GGTCTTTCT ATCTCTGGT TAGTGCATTA TCTCTGGTG 360
 50 CTCCTTACC CAGGACTTTG GGAATGGTCT TTTTGTAAATA CATTCTCTC AAATAATTCA 420
 APTTGTAGT TCTCTATCT ATCTCTCTG GAGCTCTTA TATACAAAT ACTCTGCTG 480
 AATCAATAAA TCTGTAGTA TACAAAGAT GTTCTTTT TCAATAGTT ATAATGATA 540
 60

400

TGTTCCTAAAT AACTCCMACA AGGAARTIAG CACATTTGGA ATATCANTAT CTTTCATTA 780
 TAATATCTTT CCMYGGAAA AWAATGATAT TCCMAACTGG GAGTGTCCGW ACCARATCTG 840
 5 ANTCTGTGTA TTGCCCCG GGTGGGCCAG CCCCCTAGAC TGTATGCTCT CATTCTCTTT 900
 GTTTACAAAA TTGAGATAAG CCCCATTCTT CCCCCACCC CCCCCATCCA TATTSTTTTG 960
 10 AGAATAAAAT GAGAGGATGT GTGTCAAGG TGTATTTTGG CATTAGTCTC TGAGGCATTT 1020
 TGTGAGCACC TGCATACTGT TGACACTCAA GTAATATTTT ATCAGCATTC CATTTCAGGT 1080
 CTTCCCTTAA TGAGGTGTGT GATGTACAAG AGTGTGAGG TGSCAAAGGA TGGGCTCTG 1140
 15 AGGAAACACT TAGGAAACTG GCGTTTGTGC CATTAAAGA GACAAACCTT TGTGGTGACC 1200
 TAATTAAAGT TTITAAAIT CAATTTGGA AGTTAGCAAG CTAGCTCCTK TCCAGGWAAA 1260
 20 ATAAGSAGTC AGTSCATGAC CTAACGGTTC CCGGCTGCT TGCCATTCCA AACAACTCCA 1320
 GTAAGTTTAT CACNTCTTT CAGGCACTCA GGTTCACAG CACAGACTTG GATAAGGAAG 1380
 GATGTCTTAT GGGTCACAT TGATC 1405
 25

(2) INFORMATION FOR SEQ ID NO: 150:
 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2890 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

TTATATGCTA CAGCTACACT AATTTCTTCT CCAAGCACAG AGGACTTTT CCAGGATCAG 60
 40 GGGGATCGGG CGTCACTTGA TGCTGCTGAC AGTGTGTGTG GGAGCTGGAC GTCATGCTCA 120
 AGTGGCTCCC ATGATAATAT ACAGACGATC CAGCACAGA GAAGCTGGGA GACTCTTCCA 180
 45 TTCCGGGATA CTCACTTTGA TTATTCAGAG GATCTGACG GTTTATGGGC ATCAAGCAGC 240
 CATATGGAAC AAATTATGTT TCTGATCAT AGCACAAAT ATAACAGGA AATCAAAGT 300
 AGAGAGAGCC TTGAACAAC CCACTCCCA GCAAGCTGG CTTCTTCCAC AGGTTACTGG 360
 50 GGAGAAGACT CAGAAGTGA CACAGGCATA ATAAAGCGA GGGGTGGAAA GGATGTTTCC 420
 ATTGAAGCDB AAAGCAGTAG CCTAACCTT GTGACTACGG AAGAAACCA GCCTGTCCCC 480
 55 ATGCTGTCCC ACATAGCTGT GGCATCAAT ACTACAAA GGCCTATTGC ACGAAAGGAG 540
 GGCAGGTATC GAGAGCCCC GCGACCCCT CCGGGCTACA TTGGAATTCT CATTACTGAC 600
 TTTCCAGAAG GGCCTCCCA TCCAGCCAGG AAACCGCCG ACTACAACGT GGCCTTTCAG 660
 60

	AGATGCGGGA TGGTGGCAGG ATCCTCGGAC ACAGGTGGGC CTTCATCCGT ACAGCAGCCA	720
	CATGGGCATC CCACGAGCAG CAGGCGTGTG AACAAAGCTC AGTGGCATAA AYCGAAGSAG	780
5	TCTGACCCCG GCTCGGCCCC YTATCAGTCC CAAGGGTTTT CCACCGAGGA GGATGAAGAT	840
	GAACAAGTTT CTGCTGTTTT AGGCACAGAC TTTTCTGSA GCAGAGCSAG CCACCTGAAA	900
	GGAGAGCACA AGAAGACCTC CTGAGCATTG GAGCTTGSA ACTCACATTC TGAGGACGGT	960
10	GGACGAGTTT GCCTCGTTCC CTGCTTAA AGCAGCATGG GGSTTCTCT CCCCTTCTC	1020
	CTTTCCTCTT TGCATCTGAA ATACTGTGAA GAAATTGCCC TGGCCTTTT CAGACTTGT	1080
15	TGCTTGAAT GCACAGTGA GCAATCTTCG AGCTCCACT GTTGTCTCCT GGCACATCAC	1140
	ACAGTATCAT TCCAAATTC AAGATCATCA CAACAGATG ATTCACTCTG GCTGCACTTC	1200
	TCAATGCTG GAAGGATTTT TTTTAATCTT CTTTCTAGAT TTCAATCCAG TCCTAGCACT	1260
20	TGATCTCATT GGGATAATGA GAAAAGCTAG CCATTGAAT ACTTGGAGGC TTTAAGCCAC	1320
	CAAGGAAGAC AAAGAAAAAC AATGAAATCC TTGAGTACA GTGCTTGTCC ACTTGTCTAC	1380
25	AATGTCTCTC TTTTAAAAAA AAAAAATGA GTTTAAAGAT TTTGTTGAGA GAGTAAATAT	1440
	ATATGCAATT AATGATTACA GTATTATTTT AAACCTTAAG TAGGTTGCC AGCTGTCTT	1500
	CTGAAAAACC AAATATGCCG GACAGGGTGT CGGCACACCA AGAAGACGGG AAGACCTGGC	1560
30	TTTGACCCCT CGCTTCCAT GTCCTTCTGG TCTCAGCGGC GAAGTCCCT ATCCTGSAAG	1620
	TAAGAAAGT TAGCCAATTA ATAGCAAGAC ACCTCATCTG CTCCTCCCC AGTGGATGGG	1680
35	GTTCTTCTGT AAAACTSTTT GCACATGGCC AGGGAGGGA ACTAGGACCC TTGTGTCTTG	1740
	TCTGAGCCTT ATGGAGGAG GAGCTGTCA TTGGGGATG TGTCTCTCTC CATTGAGATG	1800
	GATGCAAAAC CCCATTTTTA AGTATATTTT CTTTATTTT TGTAAATTTA GAGGTGTAGG	1860
40	TTTTGTTTTT TTTTTTTTTG TTTTTTTTTA AGAGAAACAT TTATAACTGG ATAGCAATGC	1920
	AGTGAAAGCA GTTTGGGATG TTGGAGCTAA TGCAGCTGT TTATACTGCT CTTTCAAGAC	1980
45	AGCTCCCTT TATTGAATTG GCATTACCGA ATAACAAGC CTTTAAAGCT GATAAAGAT	2040
	CAAAAAGCTG GTTAGACATG CCAGCCTTTG CAAGGACGT TATCAGCAA AACTTAACCT	2100
	CCAAGTGCTT TTATGGACGC TGCATATAGA GAAGGCTTAA GTTAGCAAC CATCTACTCA	2160
50	CAGCTGCTAT TAAGCTATA ATGACTGAAA TGACCTCTCC ACTTATTTTT TGTGTGTIT	2220
	TGCACAGACT CGGAAAAAGT GAAAGCTGCC AATCTGAGTA GTATCAAAAT GTGAGGAAT	2280
60	TTTCTCTTAA TTTAATTAAT GTTCTCTTAA TTTAATTAAT TTTAATTAAT TTTAATTAAT	2340

TGTCAATTAT GCATTTGTAA TTITACATGT AATATGCATT ATTTGCCAGT TTTATTATAT 2520
 AGGCTATGGA CTTTATGTGG ATATAGAAAG ACAGAAATCT AGCTCTAGCA CAAGTTGCAC 2580
 5 AAATGTTATC TAAGCATTAA GTAATTGTAG AACATAGGAC TGCTAATCTC AGTTGGCTCT 2640
 GTGATGTCAA GTGAGAAATG TACAATTAAC TGGTGAATTC CTCATACTTT TGATACTACT 2700
 TGTACCTGTA TGTCTTTTAC AAAGACATTG GTGGAGTCTG TATCCCTTTT GTATTTTAA 2760
 10 TACAATAATT GTACATATTC GTTATATTTT TGTGTAAGAT GGTAAGAAATG TACTATGTTT 2820
 ATGCTTCTAC ATCCAGTTTG TACAAGCTGG AAAATAAATA AATATAACAT AAAAAAAAAA 2880
 15 AAAAAAAAAA 2890

20 (2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2399 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

30 GAACTTTTCC ATCTGGCAA CCGGAAACTC CATCCCCATT AAACCAACTC CCCCTTTTGG 60
 TTTCCCCCCC AGNGGAATAG AATTTGGACN CCCATATAAA TCCAGGAAAC CACCTAAATT 120
 CTTTAGTNGT TTGTGTTTCC AAGATCTAAG GTGATGGTAA ACATTAAGTT CTAAAAATTT 130
 35 TTGGGAGGGA CCAGTGCACC TCTCCCTCTG AATGTTCNC CAATTTAAAA TTGAGTAAG 240
 GTTTTAAAAAT GTCTTATTC ATTTGGAAGGG TTTGTTATTT CATTTTGAGC CCAGAGGGGA 300
 40 GAGGCACATT TTAAATATCA GAATTAGATT AGCTTTGAGT TTGTACAATT GCGAACATAA 360
 TAGATTTTCA TAAATTATGT GTGCCTTGTT GGAAGTGCA ACTGTCTTTA TGTCTGCTTG 420
 TAAAAGTTTC AAAATATGTT TCCCTCAA AAGGCAACGT TACTTCATTT GCTTGAATAT 480
 45 TATGATAGGA ATGCTTACTG ATATTACTTG ATAGTCATAT ATAGCCTAGG AAATTTAACA 540
 TATATATAAC TATAGCAGTA TTAATAATGA TAGTTGTACT TCTTTAAAAC ATTAAATTTG 600
 50 AGGAAACTTT AATGCTGTCT CGTGACATT GCTTTACTAC AGTGAGGGGG AATATCCTTT 660
 AGATTGAGCC TCAATTTACT GGTAGTAGT ATGTGAACTC TGGTATAAAA ACGTAAACTA 720
 GACAGTAGAG CCGATGAATT AAAATTGTAA ATTGCTACAT TGGCATTTTC TACCTCCTTT 780
 55 TCTGTCAGAG TATTACTTTT TCCAGCATTT ATTCTTATTT GTGAGTAAAG AGGAAATGGG 840
 AACCTGAGGT TAAAATTGAC ATTTTGTGTT CATGAGAAT TTAAGCAGTA GGTACAGGAG 900
 60 AAGTGACTTG TCACATTAAT TTGGTGCCTA AATCTGTAAC TACAAGTTGT GATCGACATG 960

	TACAAAATGT CTAAGAAAGG TCATATGCTG AATATTTTAC TTTTCCTGTA TAGTCTGCAT	1020
	GATTTGTMTG ATAAACCCAG CTTATTTTCT CCAAAAAGCA AAATGGTCTT GTAATTTTTA	1080
5	AAGTAAAATA AACGTGCDAT TTTGTCTGCA ATCTATAATT TCAGGAAGTT ATTGRAAGTT	1140
	CTGACTCAGG GCTTTTAAAC AGTTCAAGCA ATTGTGAGTT ATATTTTGGG AACTCCATCT	1200
10	GTETAATTCT CCAGTGCTTT GAAAGAATTA TTAAGTGGC AACACTATTA AAACCTTATA	1260
	AAAGATGGTC TTTAGTGCAC GTGTATCATT ATATACAGCT TTAAAGTCA TATTGCTTAG	1320
	CTGTTAATA AAGATTCTGC ATGTGTGCTG GGTTCGGTA ATTCTTTAAA GGAAGTTTTC	1380
15	TAGATTTGCA CTTGATGTTT GTTTTAAAA AACTGATTAT TTATGGCCGT GACACTGTTA	1440
	CCAGAAAAGT AATTCTAATT AAGTTATTAT GCAAAGTCAT CTATAAGTAG CATCTGGGAA	1500
20	GAGGAGATSG AGGCAACACT TTCTATTTT AGTATGAAAG GAGGATCTGT TTGGGAAACA	1560
	TAGATTGTCT TCCCTCAAA TGAGGGGAAA AAAAAAGACC CTTTGTTCAT ATGATTTCTG	1620
	TTGTAAAAAA TTAATTTTAA AGGAAATCAC AAATTGTATG TCATTCTTAA TGCTAGTCTT	1680
25	ATAGAATAAA TCCATAAAAT TGTTTTTATG TTCAGTATGT TTATGTCATT CTAAATGCAG	1740
	CAAAATTCAT GATAGCAGTT CAATTGACTC ATAGCAGTGT TTTGTATTTT TTCTAATTCT	1800
30	TTAGCTTTCA ATATTGGATT AAAGTCTTGT TTGTGAATAT AGTTTCCGTA TGCCAAATGA	1860
	TTCTTTGCTT ATTAGCTTTT GTTAAAGAAT GCTTAGTAAG AGCTAAGCTT TTAAGAGTAA	1920
	TGCAACATT TATCGTTAAT AAAACCTATG GTGTAATATC ATATAATGCT TTTCTTTGAT	1980
35	CTTTGGAGAA TTATTTCTTT ATAGTAGTAT ACATGAATTT TGATTTTAA AGCATTTAA	2040
	AACAAATCTC AATACATTAA AAAACCTGTT ATTGTAAAA RGGAAATTAC CATGCCTTTA	2100
40	AGAAACAAGG ATGTACATCT TCAATTCAGC ATRAGTGTCC ACATCTAGAA GGCTCTCAT	2160
	GCAGTTGTTT ACAGTTAAGG TACCTCTATC TAAAGGGCCA AAGAAGCATT TCATAYTTTA	2220
	ACACCTCACA TTCTTTTCAAG ATTAAGACAT ATGAAAATAG TCTGAATAAG ATAAATTTGG	2280
45	ATAGGAAGTA ACTTAACAG TCTGGGAAGA TTCAGGCTTT TTCTATKAAA AAGCTTATTC	2340
	CTCTTACAA CTGNGGIGGT ACGNTTTCAT TTTTCAAGAG GGTAGATATT TTAAAGCCA	2399
50		

(2) INFORMATION FOR SEQ ID NO: 152:

404

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

	CCTGCGTGTA GTAAAGTCAT CCGTGCCTTT GAGATGGTGA TGGTGCCAA GSACAATGTT	60
5	TACCACCTGG ACTGCTTTGC ATGTCAGCTT TGTAAATCAG GATTNTGTGT TGSAGACAAA	120
	TTTTTCTAA AGAATAAAT GATCCTTTGC CAPACCGACT ACAGGSAAGG TTTAATGAAA	180
10	GAAGGTTATG CACCCCMGGT TGGCTGATCT ATCAACATCA CCGCATTAAG AATAAAGC	240
	ACTACATCT TTTATCTTTT TGGCTCACA TGTACATAAG AATTGACACA GGAATCTACT	300
	GAATAGCGTA GATATAGGAA GGCAGGATGG TTATATGGAA TAAAAGGGGG ACTGCATCTG	360
15	TATGTAAGCA AATTGCCCCA GTTCAGAGTT GAATGTTTAT TATTAAAGAA AAAASTAATG	420
	TACATAAGGC TGGATTTTTT TGCTTCTAT TCGTTTTGT GTCACTTGGC ATGAGATGTT	480
20	TATTTTGGAC TATTGTATAT AATGTATTGT AATATTTGAA GCACAAATGT AATACAGTTT	540
	TATCTGTATA CCATTTGTGT TCCATTTGCT YCTTTGTATT GTTGCATTTA GTACAATCAG	600
	TGTTTAAACT TACTGTATAT TTATGCTTTC TGTATTTACC AGCTATTTTA AATGAGCTGT	660
25	AACTTTCTAG TAAAGATITG AAAAGCAAAT CCTCACTAAA GGATACACAG GATAGGATAA	720
	AGCCAAGTCN CATCAACATT AAAAAATACT AAAANANAAA ACACAAAAAA AAAAAANCCC	780
30	GGGGGGGGGC CGGAATCCAT TC	802

(2) INFORMATION FOR SEQ ID NO: 153:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 461 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

45	CTAGSAGCAC CGAGCAGCTT GGCTAAAAGT AAGGGTGTGG TGCTGATGGC CCGTGCGCA	60
	CTGACCCGGG CTCTGCNCTC TGTGAACCTG GCGCCCCCGA CCGTGCCGC CCGTGCCCCG	120
	AGTCTGTTC CCGCCGCCCA GATGATGAAC AATGGCCTTC TCCAACAGCC CTCTGCCTTG	180
50	ATGTTGCTCC CCGGCGGCC AGTTCTTACT TCTGTGGCC TTAATGCCAA CTTTGTGTCC	240
	TGGAAGAGTC GTACCAAGTA CACCATTACA CCACTGAAGA TGAGGAAGTC TGGGGCCGA	300
55	GACACACAG GTGGGAACAA CGACAGGGG ATTTAAGCAG TCAAAAGGAA AAACATGTTA	360
	AGACCCCTAG CTTGTATATT GACACACTTG TACCTTGTA GGCAGAGGAA TGTAATTAAA	420
60	AAGCACTTAT TTGGCWNAAA AAAAAAAAAA AAAAAAAAAA C	461

(2) INFORMATION FOR SEQ ID NO: 154:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2388 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

10

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

GCGCAGCTGT CCGAAAGCGG AGAACCCTGG TGGGCTGTT GTGGASTAGG CTTTGGACTG 60
 15 AGAAGCATCG AGGCTATAGG ACBCAGCTGT TGGATGAGG GCCCAGGGGG GCTGGTGGCT 120
 AACCGAGGGG GCGGCTTDAAG GTGGGCAATT GAGCTAAGCG GGCTGGAGG AGGTCAGCAGG 180
 GGTGGAAGTG ACCGCGGCGG TGCGCAGGGA GACTCGCTCT ACCCAGTGGG TTAATTGGAG 240
 20 AAGCAAGTGG CTGATACCAAG CGTGCAAGAG ACAGACCGGA TCCTGGTGGA GAAACGCTGG 300
 TGGGACATGG CCTTGGGTCC CCTCAACAG ATTCCCATGA ATCTCTTCAT CATGTACATG 360
 25 GCAGGCAATA CTATCTCAT CTTCCCTACT ATGATGGTGT GTATGATGGC CTGGCGACCT 420
 ATTCAGGCAC TTATGECAT TTCAGCCACT TTCAAGATGT TAGAAAGTTC AAGCCAGAAG 480
 TTTCTTCAGG GTTTGGTCTA TCTCATTGGG AACCTGATGG GTTTGGCATT GGCTGTTTAC 540
 30 AAGTGCCAGT CCATGGGACT GTTACCTACA CATGCATGG ATTGGTTAGC CTTCAATTGAG 600
 CCGCTGAGA GAATGGAGTT CAGTGGTGA GGAATGCTTT TGTGAACATG AGAAAGCAGC 660
 35 GCTGGTCCC CATGTATTTG GGTCTATTT ACATCCTCT TTAAGCCCAG TGGCTCCTCA 720
 GATACTCTT AAATAATCA CTATGTATA AAAGAACCA AAGACTCTTT TCTCCATGCT 780
 GCGTGCAGG GTCTAGAAG GACAATGTG ATATTACGAC AAACACAAAG AAATAATACC 840
 40 ATAACCAAG GCTGAAAATA ATGTAGAAAA CTTTATTTTT GTTCCAGTA CAGAGCAAAA 900
 GACACAAAA AAAACATAAC TATGTAAAG AGAGAATAA TGCTGCTAAA TCAAGAAGTG 960
 45 TTGCAGCATC TCTTTCAAT AAATTAATG GTTGAGAAG ATGCATAAAA AAAGTTGCAC 1020
 AAGTTCTTA TTTTCTTAA TATTTCACTT CTATTTAATA CAAGCTGGGA CATAAAAATT 1080
 CTGTTGGGGA TACCTGGGGG AAGATGTGAG AAATAATGC TGAATTCAGC TTATACATGA 1140
 50 TGAAGAGAAA AACCAACAA AAGGAGCACA TAAATATGA TACAGTGTAA CTGTTATTAT 1200
 TTTAATAAGG AGAATAGGG ATTATGTTA GATGCTTAA GAGGAACGAG GATTCGTGGG 1260

 ATAATATTT AAATCAAGG GTCAAGAAA GATCTATAT CTGTAATTA TTTTCTTAAA 1320
 60

5 GTCATAGGTA AGACTCAAAA GCGGGATCTT ATTCAAAAGG CAGGTATTTT CTTTGTTTTC 1500
 TGTCTTGAAA TAGGCTCTTC CCCTAAGGTG CATTCTCTCA AGTTTTCAGT ATTGCTTTAT 1560
 10 TTGCACTGAT TAAAAGAGAT GAGAGACTTT GGAGACAGAC AACGTAAGCA ACACATACAC 1620
 ACATGAAATA CTCTAGACAG AGATGAATAT AAATCTGGCC TAATAACCAG TTTTCCATGT 1680
 AACAGTGATT TTGTGTTTCG GGCTGAAGCA GTGGTTATAT TAAAAGCCAC TAATTCCTTT 1740
 15 ATCCCTTTAA AAGATTTTTA CAATCTCCA ACCACAAACA GCACTTCTAA AACTAATTTT 1800
 ACTTTCTGCC CATAATTTGT TCTACATGGA AAAAAAAAT ATTACTTTGG CCAGGGGTGT 1860
 GTGTAAATGT GGCAGAATTC CTAGGCAGGC TGACCTTTAC AGTATGGGCC TTAAAGATAC 1920
 TGGATCCTGG TTGGCAACA AGTGTACCGT CTGAAGTTTC TGAAAACAAA TTAGAAGACT 1980
 20 GTTGGCTTGG CTAATCTCGT AGTTCAGGGC CAATTTCTG TAGTCAGAAT GAAGAATAAA 2040
 ATTGAAAGAA AAAGGGGAAA ATGCTTATAC TTGCATTAA GTTGAATGCC TCAAGTCTTA 2100
 ACTATGGCTT TGTATATGAG CCAAAAGATT TCTTAGTGGT AAAATTTCTT CAACAGGTCA 2160
 25 ATGCCAATCT GTATGCCATT TTAGTAAAGT AGSTAAGGAG AGTAGCCGCT CAGTAACTTT 2220
 GGCCTAAAG AAAGAGTGTG GTCTAGAAC TTCCAATCCC ATTGCTAGAT GTGCCCTTTA 2280
 AAAGATGGTC CAGTGGTTTC AGGGAAGGAT GTTTAGCCAG TTTTCTAGT ATTTGTTTCT 2340
 30 TAAGATTTIT TGACCTGTGC TTAATAAGAC GGACGCGTGG GTGACCCC 2388

35

(2) INFORMATION FOR SEQ ID NO: 155:

(1) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 642 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

45

AAAACAGACC ATTTAAAAAC TCAGACAAGA TTATATTAA TATATTAATT ACTAAAAAGG 60
 CACAAGATTA CACTGAACAT ATTAGCTACT AAAAAGGCAC TGCTAAGACA TTCAAGCAAA 120
 50 TAGCTATTAC ACACTACTGC AGATTTTACA GGTTCCTAAT TCTAACATAT GTTTGAAAAA 180
 TCCGTGAGTA TTCCAAAATA TATTTAATAA TGAATATCT GCATTAATAT ACCATCCATG 240
 TGTTTTACC ATTTGCCCTTA ATATTGAATA TACTGTTTAC CTCACACTAA AAAGAAAACC 300
 55 AGAAGCCTTA TTTGTGATTT TGGGAGTGGG AGCTTCCATT TTTGTGTCAA AAATGAATCC 360
 TGATTCTTAT GGAAATCTCT GTTATTAAGA TATTTCAAGA TGAGACAACA CTGAAGATCA 420
 60 AATGTGTTT AGTATCACTA TCTTCTCTCC TCGTTTCTCT CTTACTCCTC ATCCTCCAG 480

AATCTACCAG TTTATGCTAG AAAGATGGGA ACCTTATTTG AATGTGTTTT TTTTTTTTCA 540
 TGATGTCCAA TTTTGTGTG CGAAAGGATT TGGATAAAAT TTTTGTTTAA ATTTTCGTAG 600
 5 ATTTTATCT ATACAAATTT AAATAAAATP ATGTTTTGTA AG 642

10

(i) INFORMATION FOR SEQ ID NO: 156:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1251 base pairs
 15 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:
 20

GCGGCTGCCC CTCACGGAG TTGCTGATCA TCTGGGCTGT CATCCACAAA CCGGTTCTT 50
 TGTCCTCCT AATATCAAAC AGTGATTCG CTTCCTCAG AGGCGAACT GCACGTTTAA 120
 25 AGAGAAAATA TCACGGGCG CTTTCCACAA TGCAGTTGCT GTAATCATCT ACAATAATA 180
 ATCCAAAGAG GAGCCAGTTA CCATGACTCA TCCAGGCACT GAGCATATTA TTGTGTCTAT 240
 GATAACAGAA TTGACGGTA AGGATATTTT GAGTTATCTG GAGAAAAACA TCTCTGTACA 300
 30 AATACAAATA GCTGTTGGA CTCGAATGCC ACCGAAGAAC TTCAGCCGTG GCTCTCTAGT 360
 CTTCGTCTCA ATATCTTTTA TTGTTTGTAT GATTATTTCT TCACCATGGC TCATATTCTA 420
 35 CTTCATTCAG AAGATCAGGT ACACAAATGC ACCGACAGG AACGAGGTC GTCTCGAGA 480
 TGCAGCCCAAG AAGGCATCA GTAAATTSAC AACCAGGACA GTAAAGAAGG GTGACAAGGA 540
 AACTGACCCA GAGTTTGATC ATTGTGCAGT CTCATAGAG AGCTATAAGC AGAATGATGT 600
 40 CGTCCGAATT CTCCTCTGA AGCATGTTTT CCACAAATCC TCGTGGATC CCTGGTTAG 660
 TGAACATTGT AACTATCTTA TGTCCAAAT TAATATATTC AAGGACCTGG GAATTTTGGC 720
 45 GAGTTTCCCA TGTATCTATA ACGTACGATT CTATATGGAA AAGCTCACCA GAACCCAAGC 780
 TGTTAACCGA AGATACGCC TCGGAGACCT GCGCGGCGAC AACTCCCTTG CCTTTCAGCC 840
 ACTTCGAACT TCGGAGATCT CACCTCTTCC TCAGGATGGG GAGCTCACTC CGAGAACAGG 900
 50 AGAAATCAAC ATTGCAGTAA CAAAAGAAAT GTTTATTATT GCCAGTTTTG CCTCTCTCAG 960
 TCGCTCACA CTCTCTACA TGATCATCAG ACCACAGCT AGCTTGAATG CTAATGAGGT 1020
 TTTATTTTTT TAAAAATTC TAAATATTA AGTTTCTT TAAAAAAT AATAATATA 12

60

ATAAAAAAAAAA AAAAACCCCCG GGGGGGGGGG GGTCGCAAT TGGCCCTATG G

1251

5

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2157 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

15

CCGGCCGGAG AGGGAAGCTG CAGGAGAGG CGCGGATCTC AGCGGGGAG CAGTGCTTCT 60

CGGGCAGGCG CCTGAGGGAG GGAGCTGTCA GCCAGGGAAA ACCGAGAACA CCATCACCAT 120

20

GACAACCACT CAGGAGGCTC AGGACAGATA CAAAGCTGTC TGGCTTATCT TCTTCATGCT 180

GGGTCTGGGA AGGCTGTCTC CGTGGAMTTT TTTCATGAAG GCCACTCAGT ATTTCACAAA 240

25

CCGGCTGGAG ATGTCCGAGA ATGTGTCTTT GGTCACTGCT GAAGTGAGCA AGGAGGCCCA 300

GGGGTCAGCG CAGGCTGAG CAGCCTTGCC TGAGCGGAAC TCTCTCAGTG CCATCTTCAA 360

CAAGTCAATG AGGCTATGTC CCATGCTGCC CCTGCTGTTA TTCACCTACC TCAACTCCTT 420

30

CCTGCATCAG AGGATCCCCC AGTCCGTACG GATCCTGGGC AGCCTGGTGG CCATCCTGCT 480

GGTGTCTTCTG ATCACTGCCA TCTCTGTGAA GGTGCAGCTG GATGCTCTGC CCTTCTTTGT 540

TATCACCATG ATCAAGATCG TCTCATTAAT TTCATTGCTT GCGATCCTGC AGGGCAGCCT 600

35

GTCTGGTCTG GCTGGCCTTC TGCCTGCCAG CTRACAGGCG CCCCATCATG AGTGGCCAGG 660

GCCTAGCAGG CTCTCTTGCC TCCGTGACCA TGATCTGCCG TATTGCCAGT GGCTCGGAGC 720

40

TATCAGAAAG TGCCTTCGGC TACTTTATCA CAGCCTGTGC TGTCATCATT TTGACCATCA 780

TCTGTACCT GGGCCTGCCC CGCCTGGAAT TCTACCGCTA CTACCAGCAG CTCAAGCTTG 840

AAGGACCCGG GGAGCAGGAG ACCAAGTTGG ACCTCATTAG CAAAGGAGAG GAGCTAAGAG 900

45

CAGGCAAGA GGAATCTGGA GTTCACTCT CCAACTCTCA GCGCACTAAT GAAAGCCACT 960

CTATCAAAGC CATCTGAAA AATATCTCAG TCTTGGCTTT CTCTGTCTGC TTCATCTTCA 1020

50

CTATCACCAT TGGGATGTTT CCAGCCGTGA CTGTTGAGGT CAAGTCCAGC ATCGCAGGCA 1080

GCAGTACCTG GGAACGTTAC TTCATTCTCT TGTCTGTTT CTTTACTTTC AATATCTTTG 1140

ACTGTTGGG CCGGAGCCTC ACAGCTGTAT TCATGTGGCC TGGGAAGGAC AGCCGCTGGC 1200

55

TGCCAAGCTG GATGCTGGCC CGCTGGTGT TTGTCCACT GCTCTGCTG TGCAACATTA 1260

AGCCCGCCCG CTACCTTACT GTGCTCTCG AGCAAGATGC CTGTTTCATC TTCTTCATGG 1320

60

CTGCCTTTGC CTCTCTCAAC GGCTACCTCG CCAGCCTCTG CATGTGCTTC GGGCCCAAGA 1380

AASTGAAGGC AGCTGAGGCA GAGACCCAG AGCCATCATG GCCTTCTTCC TGTGTCTGGG 1440
 TCTGGCACTG GGGGCTGTTT TCTCCTTCTT GTTCGGGGCA ATTGTGTGAC AAAGGATGGA 1500
 5 CAGAAGGACT GCTTGCCTCC CTCCTGTCTT GCCTTCTTCC CTTTCTTCTT GCCAGGGGTG 1560
 ATCTTGAGTG CTCTGGGGT TTTTCTTCTT AACTGACTTC TGCTTTCCAC GGGGTGTGCT 1620
 10 GGGCCCGGAT CTCCAGCCCC TGGGGAGGGA CCTCTGGAC GGACAGTGGG GACATTGTGG 1680
 GTTTGGGGCT CAGAGTCGAG GGACGGGGTG TAGCCTCGGC ATTTGCTTGA GTTTCTCCAC 1740
 TCTTGCTCTT GACTGATCCC TCTTTGTGCA GGCCAGTGA GGCTCTTGGG CTGAGAGAAC 1800
 15 ACGTGTGTCT CTGTGTATGT GTCTGTGTGT CTGGTCCGT GTCTGTGAGA CTGTCTGCCT 1860
 CTCCTGGGTG GCCTACGAGC TGGGTCTGAC CGTTGTATGG TTGACCTGA TATACTCCAT 1920
 20 TCTCCCTTGC GCCTCCTCTT CTGTGTTCTT TCCATGTCCC CCTCCCAACT CCCCATGCCC 1980
 AGTTCTTACC CATCATGCAC CCTGTACAGT TGCCACGTTA CTGCCTTTTT TAAAAATATA 2040
 TTGACAGAA ACCAGGTGCC TTCAGAGGCT CTCTGATTTA AATAAACCTT TCTGTTTTTT 2100
 25 TTCTCCATGG AAAAAAAAAA AAAAAAA 2127

30

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 1625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

40

CAAAAGATCT ATAATCAGGA CATTGTTTAT GTAAGTTTGA CAANAAAAAT TCTTCCCTT 60
 TATGTGACCC CTTCCTATGA TTGCAAGACA AAATTTCCCT CTTTACCTC ATCCCTATAA 120
 45 CATGGGAGGC TGAGAAAAAT GAGGGAGAT GGAACAGAT ACAAGGAGAT CCAATAAGAG 180
 AAGCTTATTT AAATATTGTG AAATAAGGA AGAMCCAAAG CATTTTTTTA AGTGGGGAAT 240
 CCTTTTGAAC AGTTATTATT TATCCATATT ATTAAYAACA TCTTTTCTGA CAAAATCCAT 300
 50 CAGATGAAGT GTAAATGGAT AATCTTTTAA TGATGTAAA CTAGAAAGT TTCCTTACT 360
 TCTGATGATG TCTGATGATG TCTGATGATG TCTGATGATG TCTGATGATG 420

60

TTATATTGAA GGAAGTAAA TTTAATTTT TTTTCTTCTT ATATTTTAA GTTCAAAAT 480

	TTGAAGTCTC AGTCACCAGA CACAGGTTCT ATACAATTAA TGATGAGCTG GAGAAGTAAT	650
	ATGTAGCTAA TTTTTCAAAA GCATTGAATA TACTTTCCCG AAAGAAAACA GAAATTAAAT	700
5	ATTGCCACAT CTTCGCCAGAA TCCCATCTGA CACCTTAACT TTGTGAGGTT TCCACAACT	750
	TGCTAATCAA STTTTATACA TTCTAANTCT CCCCAGTTTC TTTCGGGCTG GAAGATGCAA	800
10	CTTCGATTTA ATAGAAACTT TCAATCTTG GGGTAAGGGA GCACTGGGGG GACTAGGGAG	850
	AAGGATAAGA AATAGAATTA TCAAAAAGCC CCCACCAGGG ACCTTCCTGG CCAGAATATG	900
	CAGAGTAATT CTGCGGGCT TCACTTTTGA AAGTCCCTCG AAACTATGCA GATGAAACTG	1020
15	AGTCTGTTTT TGATATTCTC AATCTATTTC TACTTTGGAA GTCCCNACAC CTAAACTGGA	1080
	ATCTCTGTAT TTATATCTCC TCCACTCTCC CCCACACCAC CCGTCAATTG CTGCTGCTCC	1140
20	TGCTAANTTT AAGCATTTTT CTCTGTTAT CATCAGGTTG ACATTAAAAA CAGTACTTTA	1200
	CAAACTGACT TCAAGCAGAG ATACTTTTAC GAATGTGATA AAATATTTTC TTAAGAAAAG	1260
	GAAAGAGGAT CTGGGTCAAA TAAAACACCG CATGGATGTT GATGCTGAA TACTGCTGTA	1320
25	AGAAAAAGGA GCTCAGCAAT TTTTATTACT CTATTTGTAA ATGAGTTTGA AGCAATTTGT	1380
	AAATGCCACT GGTACATTTT TAAGGTGACA CATTTGCTCC TTATAAAGTT ATTAAAAATT	1440
30	ACAGGCTAAG CTCAAAATGAC GTTTCACACT AGTTTTACTT TATATAATCA ATATTGATAT	1500
	TGTTGCTGAA CTATGTAAT TTATGATGCA TTTTTCAGTC CCTTTTCAGA GCAAAATGTT	1560
	TTGCAATGGT AATAATGTTT AGTTTAAATT GACTTAATAA ATMTTACCT GAGCAAAAAA	1620
35	AAAAA	1625

40 (2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1687 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

50	CGGGGTCACC AGTTATTAGA GGAAGTAACA CAAGGGGATA TGAATGCAGC AGACACATTT	50
	CTGTCCGATC TGCCAAGGGA TGATATCTAT GTGTGAGATG TTGAGGACCA CGGTGATGAC	120
	ACATCTCTGG ATAGTGACCT GGATCCAGAG GAGCTGACAG GAGTCAGGGG ACATCAGGGT	130
55	CTAAGGGGAC AAAAAGGTAT GCGACTTACT GAAGTGCAAG ATGATAAAGA GAGGAGGAG	240
	GAGGAGAATC CACTGCTGGT ACCAATGGAG GAAAAGGCAG TACTGCAGGA AGAACAAGCC	300
60	AACCTGTGGT TCTCAAAGGG CAGCTTTGCT GGGNATCGAG GACGATGCCG ATGAAGGCCC	360

411

TGGAGATCAG TCAG3000CAG CTGTTATTTT3 AGAACC03Y3 GAAGGGAC03 CAGCAGCAGC 420
 AGAA3CAGCA GGTGCCACAG ACACCC00CTT CCTSTTT3AA GACTGAGATA ATGCTC00CC 480
 5 TGTACCAAGA TGAAG000CT AAG3NAACAG AG00TTCTTC GGGGACAGAA GGTGCCACTG 540
 CCTTGAAGG GGAAGA000G GATGGCATCT CAGACAGTGA TACCACTACT AGCAKTGAGG 500
 10 AAGAAGAGAG CTGG3A00CC TCCSTGGTAA GAGCGAASC GTGGGCTAA AGTCAGATEA 560
 TGACGGGTTT GAGATAGTGC CTATTGAGGA CCCAGCGAAA CATCGGATAC TGGACCCCGA 720
 AGGCTTGTCT CTAGTGTGTG TTATTG00TC TTCCAAAAG GCCAAGAGAG ACCTCATAGA 780
 15 TAACTCCTTC AACCGGTACA CATTTAATGA GGATGAGGGG GAGCTTCCGG AGTGGTTTGT 840
 GCAAGAGGAA AAGCAGCACC GGATACGACA GTTGCCTGTT CGTAAGAAGG AGGTGGAGCA 900
 20 TTACCGGAAA CGCTGGCGAG AAATCAATGC ACGTCCCATC AAGAAGGTCG CTGAGGCTAA 960
 GGCTAGAAAAG AAAAGGAGGA TGCTGAAGAG GCTG3AGCAG ACCAGGAAGA AGGCAGAAGC 1020
 CGTGGTGAAC ACAGTGCACA TCTNCAGAAC GAGAGAAAGT GGCACAGCTG CGAAGTCTCT 1080
 25 ACAAGAAGGC TGGGCTTGCC AAGGAGAAAC GCCATGTCAC CTACGTTGTA GCCAAAAAG 1140
 GTGTGGGCTG CAAAGTG0GC CGGCCAGCTG GAGTCAGAGG TCATTTCAAG GTGGTGGACT 1200
 30 CAAGGATGAA GAAGGAC0AA AGAGCACAGC AACGTAAGGA ACAAAGAAA AAACACAAC 1260
 GGAAGTAAGC AGAGCTG0CA G0CTCCCAGG AGAGCATGGG GACTAGGAGG AAGGGTGTGG 1320
 CATGGCTCAG TCTGGCC0CC TTGATTACCG GCCTAGCCCC TGCTCACATC ACAGCTGTCT 1380
 35 GAAGAACAGT GAGGTGGAGT G0CTAGAAAT C00GTGGTGG T0CTGAGCAG AGAGGAGGAT 1440
 GTCTCCTGTC CTG0CTGAAG GTCTCCCATG AAAACACTGC TGAAGTGTGT TGACACTCAT 1500
 40 GACCCTTTTT TTAAACCGTT AAAGGGA0ST T00GTGTTGG AGCGATACTC AATGTAGTCA 1560
 GTCTACACCT GGAAGTGTGG G0CACTTAAG C0CTCCCAC C0CCATCCTA TTCTTRAATA 1620
 AAACCAGGAT AATG0A0FAA AAAAA000A AAAAA000A G0GGGGCCCN TAAAG0GNCC 1680
 45 CANNNTT 1687

50

(2) INFORMATION FOR SEQ ID NO: 160:

(1) SEQUENCE CHARACTERISTICS:

60

XII SEQUENCE DESCRIPTION: SEQ ID NO: 160:

	GGATGACASA TTGCSACANA GATTGTGTGAC CCTTCCTGCT GAACTTCAGA GGGAGCTGAA	60
	ANCAGCGTAT GATCAAAGAC AAAGGCAGGG CGAGAACAGC ACTCACCAGC ACTCAGCCAG	120
5	CGCATCTGTG CCCCAGAGAAT CCTTTACTTC ATCTAAAGGC AGCAGTGAAA GAAAAGAAAA	180
	GAAACAAGAA GAAAAAACC ATTGGTTTAC CAAAAAGSAT TCAGAGTCCT TTGAATAACA	240
10	AGCTGCTTAA GASTCTTCA AAAACTCTGC CAGGGGCGTG TGGCAGTCC CAGAAATTAA	300
	TTGATGGGTT TCTAAACAT GAAGGACCTC CTGCAGAGAA ACCCCTGSA GAACCTCTCTG	360
	CTTCTACTTC AGTGTGCCA GCGCTTTCTA GTTTGAGTTC TGACCCAGCT GCGTGTGTGA	420
15	GACCTCAGC ACCAATCTA GCTGGAGCTG TTGAATTCAA TGATGTGAAG ACCTTCTCA	480
	GAGAATGGAT AACTACAATT TCAGATCCAA TGAAGAAGA CATTCTCCAA GTTGTGAAAT	540
20	ACTGTACTSA TCTAATAGAA GAAAAAGATT TGGAAAACT GATCTAGTT ATAAATACA	600
	TGAAAAAGCT GATGCAGCAA TGGTGAAT CGGTTTNGAA TATGECATTT GACTTTATTC	660
	TTGACAATGT CCAGGTGGTT TTACAACAAA CTTATGGAAG CACATTAAAA GTTACATAAA	720
25	TATTACGAGA GAGCCTSAIS CTCCTTGATA GCTGTGCCAT AAGTCTTGT GAGGTATTTG	780
	CAAAGTCCAT GATAGTAATG CTCGGAGTTT TTATAATTTT AAATTTCTTT TAAAGCAAGT	840
30	GTTTTCTACA TTTCTTTTCA AAAAGTGCCA AATTGTCTAG TATTGCATGT AAATAATTGT	900
	GTTAATTATT TTAGTGTAGE ATAGATTCTA TTTACAAAAT GTTTCTTTAT AAAGTCTTAT	960
	GGATTTTAC AGTGAAGTGT TTACAGTTGT TTAATAAAGA ACTGTATGTA TATTTGTAC	1020
35	RGGCTCTTTT TGSTBAAYCC TTA AAAACTC AACTCTAGGA RGCACCTAAT GTTTACTATA	1080
	CTAAARGGCT GAAAAMCTC CAGGCCAGAC TGCTAAGTTC TGAATYCCT GAGAGTCTC	1140
40	AGACGGGAT TCTACTTGT CCAAGAAAGG GTAAAGCTTC TAAACCATCT TATTCTTCTC	1200
	TCCAAGCATG AACACAGGAG CATGTYAAGA AAATCTTTAC TACTTTCTYC CATGCCGAGA	1260
	AATCTACATA TTTTGAATTA GAAACACCCT CACACCCACT TGAAGATTTT TTTCTGGGA	1320
45	ACATTATCTC CCGTAGATCA GAGGTGGTGT TGTCTTTTTG CTTCTACTGG CCATTGAGAA	1380
	ACTTTGATSA TAAAAAGAA CGGTATAGAT TTTTCAAACG TATATAAAAT ATTTTATGT	1440
50	TATATGTTAT GCCATAACTT TAAAATAAAA ATAGTTTAAA ACTCTATGCT AGTGGATATT	1500
	TGGAACCTTTT TCTCAAACA AACACCCAC ACTGACTTCA GAAAAACCTT AAAACTAGCT	1560
	ACAGATTACT ACTACGAATG AATCATYAAG TTTTGTGTCT GCAACAATTT AGAAGCACTA	1620
55	AGCCCAAATA TCACGAAATG TGTGTATGAT GGAATTTTCT ACGACAAAAC AGATCAAGAT	1680
	TAAAACAGGA TCAAGGATTA ATGGTATAAA AATGGTCTAC TAAAACAGGA TCAAGGATTA	1740
60	AAACAGGATC AAGGATTAAT GGTATAAAAA TCTCTACTGG TTACGGGGTG GCGGCGCCAT	1800

ACAGGGTAGT GGTGGATGGA TAGTTTAGTT TGGNAAGGGT AA

1842

5

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 770 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

15

GGCACCAGCC CTATGCTGTT CTTGTGATAA TGAGTGAGTC TCACAAGATC TGGTGGTGT

60

ATAGGCATCT GGCATTTCCC CTGCTGACCC TCATTCTCTA TCCTGCCACC CTGGGAAGAA

120

20

GTGTCTTCTG TCATGATGT AAGTTTCTG AGGCCTCCCC AGCTATGTAG AACTGTGAGC

180

CAATTAAACC TCTTTTCTCT ATAAATTATC CAGTCTTATA TATTCTTCA TAGCAGTGTG

240

25

AGAACAGATA ATACCGTAAA TTGGTATCAC AGAGAGTGGG GTGTTGCTAT AAACACATCT

300

GAAAATGTTA AAGCAAATTT GGAAGTGGT AACAGGCAAA GGCTGGAACA GTTKGAAGAA

360

CAGTTAAGAA GAAGACAGGA AAATATGAGA AATCTTGAAA CTTCTAGAG TCTTAAAGGT

420

30

CTCAGAAGAC ATGAAGATGT GGAAGCTTT GGAAGTTCCT AGAGACTTGT TTGAATGGCT

480

TTGACCAAAA TGCTGATAGT GATATGGACA ATGAAGTCCA GGCTGAGCTT ATCCAGACAG

540

35

ACATAAGAAG CTCGCTGGGA ACTTGAGTAA AGATCACTCT TGCTAGGCAA AGAGACTGGT

600

GGCCTTTTTT CCTCTGCCCT AAGATCTGT GGAAATCTGA ACCTGAGAGA CATGATTTAG

660

GGTATCTGGC AGAAGAAATA TCTAAGCGGC AAAACCTTCM AGAGGAAGCA GAGCATAAAC

720

40

GTTTGAAAAA TTTGCAGCCT GACNATGGGA GACCAAAGTT AAACCCAATT

770

45

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 519 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

60

AGTGTGAGA AATGCTCTT TGTGATGTTT AATGCTCTT AATGCTCTT AATGCTCTT

190

414

ATGSAAGTST TTGTSAAGST TAAATGGGAA GACATAAAGC ACTTAGCCCA GATCCCAAGGA 240
 CATGCTGAAT AGGATAATGG TGGCTCTCTT TGGCGCTGTG CTGCTGCAGG TGTGCCGAGG 300
 5 AAYTGGGCAG GGGTGACAGA TACCTCTTCT AACCTAGTTC CTTTCCAAGA ACCTAATTGG 360
 TGTCTCTCCC TCTCCAGGC AATTGGAAGG AGGAGGCTGG GCGCCAGCCG CAGAATACGG 420
 10 GAGGTTTCTC ACCGTGGTAG GGAAATTGCT GGGTTGGGGG TGTGGGCAAC CAGGTGATC 480
 GTCTCTTTCG AGGADGGATG AGGCTTTGCT GACAGAGGC 519

15

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 753 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

25

GGCAAGAGGG GGAAGAGCAG CCACTTCTTG ACTGGCAGAT GGGCTCCAGC GTCCCGGCTG 60
 GTGGGACAC TAAAGCCGGA GCGATCTTCT TAATTGTAA ATTGGATCTT GAAGCTTCAC 120
 30 TGTTTAAATC TTTTCAGTGG CTTCCTTTG TACTTAGAAA AAAATGCAAC TTCTTCTGCT 180
 GGGACTGATC GGTTCACAGC CTTCCTCTCC ACCCTCTCTC TGCTTCATGC TCTGCCCCTG 240
 35 CCTGCCATGC CTGGATACT CACCTTTTGT ACCCCAGCAC CCGTGCCCTC TGCCCTCGA 300
 TCTTTGCTG GGTGGTTGCT CCACTTCAG TGTTCAGGAC AAATGCTCCT GGGCTTACCC 360
 CATCTAGCCA GTCTAGCCCG GTCTTCCTG TCTTCCCTGT TTCATTGATG GCTCTTATG 420
 40 TTTGTIWAAT TGTGTGCTGT TGACTTTTAA CTCTCTCACT CCCCCTGGA ATGCAAGCGA 480
 TCTCCCAAGC TCTAGAAAT GTTCTGCTT CTTCACAGGC CTTACGCTG TGTGTGCTCG 540
 45 TGCCGAATTC GGAAGAGGG TATGTGCACT TGCTGGTATG TATGTAGGTG TTTGCTAACA 600
 CATACGTGCA CAGGAGAAT GTTTCAGGG GACTGCACAG CCTCTAGTTC GCAGCCCCCA 660
 CCCCCTCCCTT TGGCCCTGCA CTCTCCCTC TCTGAGCTG ATTGGCATGA AAGGCTGCAN 720
 50 GGTTCCTGAN CCGGAGAGG NCACCTCCTG GGA 753

55

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1400 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

5	GGCACAGTTT ATTAATACCT ATTATGGGAA AGTCACTTTG GTTGGCATTG AAAATTACAT	60
	CATCTTTTAAA GCAGTATTTC TCCCCAGATG GACTCATCAC TAGCAAAGAC TAGGTTTCAT	120
	GGAAGGCATA GGGTAGAGA ATGGGAACAT GAGTGGAGG CGGTTTGTTA AAGTGCTGTC	180
10	AGTGAGTGAT TTTGTCTACT TGAATATGG TCCATGTTTG GGGGCATATT GTGTTTCATA	240
	AGAAGTGAAA GGTATTTTCA AAGTAAGCTA CAAATGACCC ATAAATCTGT TAACAACAGT	300
15	CCTTAATATG CAAAGATGAA AAACAGCAT TACTGCTACC CAAAGGGAAC TGSTGCTTGG	360
	TGATGTGCAG ATGGGGCTGT TGGTAAAGAG AGCTATTACA GGTTTTCTCT CTTAGGTTTC	420
	ATAGGAGGTA GTTACTGAGA TGAGATTGTT TTATCTTTTT GAATACAGAT CTCTTGCTTT	480
20	GAGTTAGTTC TGAGATGGG AGTAATAAAG GASTTTTTTG TTTTMTGTTT TGTTTGTTTG	540
	TTTTGGCTCC TTAGTAATAC TCCTCTGACA TTTATTTCTA TTATCTTCA AAGAAAGGAA	600
25	ACCAACTGAA ATGTTTGCTT TAACAAACAT TTTAATAAGT TCTCTGGGT TTTTTTCCC	660
	CTTTTAAAAA AATTAGCATA TACCATAGCA ATAAAAGAAC TAATGTAAAC TATTGTATGC	720
	TACAACITAA GTGATTTTTC TAAAGAAGCA CAATGTCATT GRAAGTATTA TTGAAAAGGA	780
30	TCATAGTCAC ATTGAATTTG TGAAGGCCAA AGAAATGAA GGGAGTGATA TTTTCATTTT	840
	ATGATATTCA CATATTTAGT AATTTTGTG TACAAGAATA CCAGGCAGAG TGTMTTACCC	900
35	ATGGAAACAG GTTTCAGATT ACTTTGTTTT TACTGTTAGA GTCTCAAGTT TAGAAATGCT	960
	AACACTTAAA TCAGTTTTTT TCTCACTATA CTTGAAGATT GTTAATATTT TSATATCTTC	1020
	CTAGCTTGAT GGAATTTAAA CATATCTTCA GATCTGTGAC AGTGACAGCC AATAGGACTG	1080
40	ATAATATTAG CTTCAAACCA ATAATATCCA GGGTTAAAT AAAAATCATA GTGAAAGTAC	1140
	GATTGTAAAA TTATGCTATA TTAACITTTA ASTCTSTAAT AACTTGACAT CAAAATGTIA	1200
45	TGTAATTAAC ATAAATAATG GCTAGCGAGA ACATCTTTTG AAATTTCTCA ATTACCTTTG	1260
	TTACTACACT GTTTCAGAAA TGAATGTAGA AATGATCCTG TTACTTTTCT GAATGTTCTG	1320
	TGGTTGAATG TGTMTTGTCT TAAATAAAGC TTTTGGTAAT TGTTTAAATW ACAAAAAAAA	1380
50	AAAAAAAAA AAAAACTCGA	1400

SEQUENCE CHARACTERISTICS

(A) LENGTH: 144 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDELINES: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

5	CAGGECTCAG GGCCTCTGGT GGCCTCTGGC CAGACAGTAT TTGAGTTCT TGTGCTATGS	60
	GTGGAGTCT TCTTCTCAA GTTCTGGCAG CTGTGCTGTG NCTGGATGGG CTGCTCCTCC	120
10	CAGGECTCAA GGCCTCTGGT GGCCTCAGG TCTCATTTCC CCAGGCCAAG TTCAAGGCAG	180
	CAGGCTTTT TGAAGGCTC TTGGCCTGG GCTGAGGGA GAACTTTAAG CTTTTTGTCT	240
15	CACAGGGAGG TGGTATGGG CCTGGGTGCA GGTGCCACCA TTCTGCTAAT GAGAGCTTTG	300
	TCTGATCACT CTTGGGTGCA TCAGTTTGTG CATGTGTCCG GCTGCCAGCC CGTCCCTTGG	360
	GATCCTTCCC CTGGGTGTA GCTTGTGTA TTASTATATA CTCATTCCTT CATGCTTTCC	420
20	TCAGCAGAAG ACTTCCACTT CTGAGGTGAG CTTTGGCCCT RTGCCCTTCC TCACAGGTG	480
	TTGCCCTTTT ATAAAGACT GATAGCAGAA TAAATTGGTG TTCCCTGTG GACCCAGCAC	540
25	CATTCTGTG GGCCTAGAAT ATGCCCTCA ACCCTTAGAG TGGGGCAGTG AGGCTTGAG	600
	GAGTGACCTT TCTTTCTCA TGTTTTAGT CATTCTGGCT GGCAGCCCTT AATGGCAGG	660
	ATCTGCTGCT TCTAACAGAT GGCAGGAGG TGACCCGAT TTCAGCCATT GCAAGGTTA	720
30	GCACCTCTC CTTTGAGCCT AGGCCACAC TGTTCAATTG CATTTTAGGC AATGCTGTG	780
	TTGGCTTTAA AGTAAGCCT GGCAGGTG AGAAGCCTG GTAAGTGATG GACTCATTTG	840
35	CTGGTCTTA AAGATGCAG CTTTAAGG CTCTTGTATG GATGCCATCT CTCTAGCCC	900
	CCAGCCTTGG TGCCACTGGT GGCAGGTT CCATTCTTTG GGCCTGGGAG GGACAGCTTG	960
	CCTGTTTCTG CTCACAAATT ACAGTCTTCT CTCTGTACT ATTCTGTGGC TTCAGCATG	1020
40	GGGCASTAGC CTTTCATTAG TGTAGATAGT CATTCCTTGG TAGGGTGGAG GGTAAAGACAT	1080
	AGGCTCTGGA ACTGTTTGGG ACCTTTTGGG GATGTCTGT GCTCCCGAGA TTCCTMGATT	1140
45	CTGGGAGGAG AGGCTGCCGC ATTCTGCTGC TCCTCAGAG GAGCAAAGCT GCACCCACTT	1200
	ACATTCACTA TTTTCTTGGC ACTACAAAGA GTGGGAAGGC CTGGGATTTG CTGCTGCTCC	1260
	CTTAGAGCAG GGCCTCTTT TTCAGCACTT TGGACACTG GAGACCCAGC CCTGTATTTT	1320
50	AATGGTAGTG GGCAAGTGTG TGTGCATACT GTCTGCCACT GCTTTCTCCC TGCCCATGC	1380
	CAGAGAGCCC TGTCCCTTGC AGGCCAGGC TTCTTAGCCC CAACTTGGGA ACAAAGTGCA	1440
55	ACATGGGATC ATGGGTGGG GTGCTCAGST GAGCCCTCTC TATAGTGCTT CCCTGGGCCA	1500
	AGCTGACACC AGCCCTGAG GGTGGGTGG GACGGGTGGT GCTTAAAGA GGAAGGGGAC	1560
	CAGTGTAGCA ACTTGCCAGG GACCCACCC CTCCCTCTCT GGGCTGTGC AGTGAGCATG	1620
60	GGGATTCCCA TCAAGGGGCC TGGCACCTGT GCTAGTTACG TAGCCGCTGN TCACGCGCTC	1680

ACTCCTGACC ACATGCACGT TCCCTAGATG CAGACTGCTT TGAACTTTAA AGCTGTACAA 1740
 TTTGTTTATG TTTGTGCTGA CTTAAAATAT ATTTTAATGA GAAAAAATA ATGGAGAACC 1800
 5 CTGGAAGGA CCTGTTCTT TGTCTCTCG GGAAGCTGA ACCCTCGCG TCTTGGGAT 1860
 GGTCTCTCG TGTCTTTCC TGAAGCTAA GCTGTCTCC ATCGCCCGAG GCTGCGCGG 1920
 10 GTCTCCCGG CGAGTTGGG TTTCTTTGG ACCTTGCTG CGGGGAGGG GTTGTCTGT 1980
 CCGAGCGCG TCTTTCTGT ACCTCTAGG CTGCGCGCG CTTTGTCTG TGAGGTCTG 2040
 TATGTCAAAA ATAAAGCGC TAGAAACGA AAAAAAAAA AAAAAAAAA AAAAAAAAA 2100
 15 AACTCGAGG GGGGGCGGT ACCAATTAA CCNNTATGA TCTATAAAG GTC 2153

20

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1251 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

30 GCGCCAGCGT CCGCCACGC GTCCGGCGGT GCGAGTATG GCGCGCTGAT GCGCATGGAG 60
 GCGTACTGGC GCTTCTTGGT GCTGCTGGG TGGCACTGC TCGTGGGTT CTTGTGGTG 120
 35 ATCTTGGGCG TGTCTGGGT CCTCCACTAC CGAGAGGGG TTGGCTGGA TAGAGCGCA 180
 CTAGAGTTTA ACTGGACCC AGTGCTCAT GTACCGGCT TGTCTTCAT CCGGGCATC 240
 40 GCGATCATG TETACAGAT GCGGTGAGC TGGAAATGA GCAAGCTCCT GATGAAATCC 300
 ATCATGCGAG GGTAAATGC AGTTGCTGCC ATTCTTGCA TTATCTCTGT GGTGGCGTG 360
 TTGAGAAAC ACAATGTAA CAATATAGC AATATGTAC GTCTGCACG CTGGTTTGA 420
 45 CTGATAGTG TATATAGTA TTTGTTACG GTTCTTTAG GTTTTCAAT CTTCTGCTT 480
 TTAGGGGTC GCTTTCTCT CCGAGCATT TCTATGCGA TACATCTTA TCTTGAATT 540
 GTCATCTTG GAACAGTAT TGGACAGCA CTTATGCGT TGACAGAAA ATGATTTTT 600
 50 TCGTGAAGG ATCTCTCAT CATTACATC CCGCCAGAAG GTGTTTGTG AAATACGCTT 660
 GCGCTTCTGA TCTGTGTGT GGGGGCGTC ATTTTGTGA TAGTCACAG ACCGCAATG 720

60

AATATGAA TATATGAA TATATGAA TATATGAA TATATGAA TATATGAA TATATGAA

ACCATGTAAA ATGTTGTAGA GATAGAGCCA TATAACGTCA CGTTTCAAAA CTAGCTCTAC 960
 AGTTTGTGCTT CTCTATTAG CCATATGATA ATTGGGCTAT GTAGTATCAA TATTTACTTT 1020
 5 AATCACAAAG GATGGTTTCT TGAAATAATT TGTATTGATT GAGGCTATG AACTGACCTG 1080
 AATTGGAAAG GATGTGACTA ATATAAATAA TAGCAGATAT AAATTTGGT TATGTTACCT 1140
 TTATCTTGTG GAGGACCACA ACATTAGCAC GGTGCCTTGT GCAKAATAGA TACTCAATAT 1200
 10 GTGAATATGT GTTACTAGT AATTAAATTGG ATAAACTGCC AGCATCCCTG A 1251

15

(2) INFORMATION FOR SEQ ID NO: 167:

(1) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 882 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

30

35

40

45

50

55

60

GACSMCTTAA AACTATGGTC CCCCGGGACT GCAAGAATTC GGCACAGCGG CTGCGGGCGC 60
 GAGGTGAGG GCGCGAGGTT CCCAGCAGGA TGCCTCGGCT CTGCAGGAAG CTGAAGTGAG 120
 AGGCCCGGAG AGGCGCCAGC CCGCCCGGG CAGGATGACC AAGGCTCGGC TGTTCGGGCT 180
 GTGGCTGGT CTGGGTGGG TGTTCATGAT CCTGCTGATC ATGCTGTACT GGGACAGCGC 240
 AGGTGGGGG CAATTCTACT TGCACAGCTC CTTCTCTAGG CCGGACAGCG GCGCGCGGCT 300
 GCGCAAGGCG GCGCGGAGCA GGGACAGGGA GCTCAGCGGC GAYTCGATG TCGAGGATTT 360
 TCTGGACAAK TTTCTCAGTG CTGGCTGAG GAGAGGTGAC YTTCTGAGAA AGGAGAGGGA 420
 GCAGCGCGCT GCGCGGAGGA GCATGGAGGA GAGCGTGAGA RGCTACGACT GGTCCCGCGG 480
 CGAMGCCCGG CGCACCCAGA CCAGGCGCGG CAGCARGCGG ANCGGAGGAR CGTGCTGCGG 540
 GGCTTTCTGG CCAAYTCCAG CTTGGCTTTC CCCACCAAGG AGCGGCGATT CRACGACATC 600
 CCCAACTGGG AGCTGAGCCA CTTGATCGTG GACGACCGGC ACGGGGCCAT CTACTGCTAC 660
 GTGCCCCAAG TGGCTGACAC CAACTGGAAG CGGCTGATGA TCGTGCTGAG CGGAAGCTGT 720
 GCACCGCGTG CGCTACCGC GACCCGYTGC GNTCCCGCGC GAGCACGTGC ACAACGCCAG 780
 CGCSCACTGA CTTCAACAAT TCTGGCGCGG CTACGGGAAG TCTCCCCCAC CTCATGAAGT 840
 CAAGCTCAAG AATACACCAA TTCTTTCTGC GCGACCCCTC TG 882

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1208 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

5 GGGAAACTCA AAAGGATGAT GGAATGGTTG ATGGAGCCAG AGCCTAGAAG TRAAGGGATA 50
 10 CAGAGTGAAG ATAGAGGTAT TTACGTATAT TTWAATATTA GTTTGGAAAT TACGTAGGGA 100
 TTCTTAAGAA AAGATCATGA CAGGACAGGC ACATTGGTA AAATGTCAGG GCAGCCAGTG 150
 15 CATGGTCTTC CTGGGGCTCC TCAGTTGAGG GGTTTAAATC ATTTCTGAT CCCCCTGCCC 240
 TGGTTTGAAG AATGCATACA GTACGTGAAA TGCCTGTGGT ATGAGTTGCA ATGGGCAATC 300
 AACCTGGGTA AATCCAAGAT TAATGATTAG TTCTAAAGAT CCAGTTGAAG TTCTAGAGTG 350
 20 GGAATTTTTC GTCAAGCAAC TCAGCACAGC TTTATGCCTG TTCTCTAAT AACGATAGGT 420
 AACAAATAGC TGTGKTWCA CAGCTAGGAR GATAACCAA TCTAGAGTTC TTGARTCTCA 480
 25 TTTAATAAAT AATATTATG AATAACAACT GCATATTTCA GGCACTGCAT TTSACTCTGT 540
 TAAATACTCA TCTTTAKGA CMSCCACWTC AGAWAACMTT AATCTGTCTG ATCAATAAAC 600
 AGCTTGACTT AGAGRGGTAA AATAGCTTCC CACAGGTWAC CCAATTAGTA GGTAACAGCG 650
 30 ACAGAATAAC AGTGCAGTTA AAATCTTABA CTGGAGACTA ATTGCATAAG TTTGAATTTT 720
 AGTTCTGCTA TGTAATTTG GGTGAGTACC TTAATTYACC TGACTCTCGG TCTTTATATC 780
 35 TGTAGAATGG AGCTAATGAT ATTACTTAAT TTGCTTTATG TGACATTAAA TGTACTAATA 840
 TATGTAAATC ACTTACAACA GCAATTGACA TATTTGACAT ACTTAATATA TTTGCTACTA 900
 ATACTATTAG CAACAGCATT CTGATTTTCC AAGTTGAAAT TCAGTGTITT CTTTTTACT 960
 40 TTGCCATAAT TTACAATGTT GTGCTCTGTA AACCATAAAT TTCCCTGAGG TGTGTGTCAGG 1020
 TTAACAAAAA ATCACTATGG CCCCARNMA CTTCGAAAAT AGAATGAGA CCAGCTTCAT 1080
 45 CTATATTCTT TACTGCAAAAT AACTTAGAAT TGTAATAGGC TAATATGTAC TGGGACTTCC 1140
 AATTGCGGAA TATGACAAAA ATANTACTAT TTAGCTAAAA CATAPACAAA ACTTATTTT 1200
 50 CCTCTGAA 1208

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5 GGCACGAGAG AAAAAGGTT GAGAATGTTT TCTAGCAGGC AGAATGTGCA TACATGTTTT 60
 CATGARTGTC CTTTGGGTGC TGTTCCTTTT AAATCCTCTG TGCACAGGGC TCTGGCCTTT 120
 ARTAAACTGT TTTTCTGTCT TACGTCATGC TGACTGGCTG CTAGGGGCTG ATTACAAAGG 180
 10 GGAAGAGTTG AACAGACATC AGGGGCCGAT GAAACCAAAG GACTAAGGAT CAGSAGAACA 240
 AGTCAGGGAT TAGGAGACAG CGSTTTGGTT TATTGTTATC CAGCTEGAGG ACTCCTAGGG 300
 GCACAGCAG GAGGAATACC AGGGCCACGG AGGGGCAGGA GTCTCACAST GGAGGCGAGA 360
 15 CTCTAACAGA TCCAGCTGA ACCTCGCTG GCCCTGGATG TCATACGAST TGGGGACCAG 420
 AAATCTGGC TCAGAGAACC CGTCCAGGGA GATTTGAAGC CATGGGTTAT CTTCTAGAGT 480
 20 TGATACTGAT AATATATTTT AATTTTATT GATGTTTAA ACCTTCTGAA ACAGGAGGGT 540
 AAGATCAGAT GGAAGCCCT TCTGTTGAAG GATCTGGGA ACCTTGGTGG TTTTMTTTT 600
 TGGTMTTTT TTTMTTGTAT CGAGCTGTGG ACATCCTTCT TAATTCGATT NTGAGGATTT 660
 25 GTTTAACTAA AAGTTCCCA AACACAGAAA GGGCCTCCCC ACCTGCTTTG GGGAGCTGTC 720
 TGTCTGGGA GTGCCAGGCA TCCSATGGGA CCCATCACTG CCAGTGTCTG TGCTCCCAG 780
 30 AGGTCAGCCC TGTCTCTGCC CTGGCTCTGT CTCTCTCTGT ACAGGGCAGA GCATTTCTGG 840
 TCAGTTTCTC CATGGTCTCT CCCACCCCTT TGTAAAGTGS ATGGACATGA TGAATTTCAG 900
 TGTCTCACC CTGATACTCT GGGTGTGAT ATTCACTTTA CCGCACTCA GACACAGGCG 960
 35 ACCTTGAAGC AGTCTCTGCT GTGTAGAGTC CACGTGACAG TCCCCACAGC CTCCCCAGAT 1020
 AGCTGTGTGC CTGTGCGCTA CTGCTGTGCC ATTTTCCCAA CTNNGGCGTT TCACTAAATG 1080
 40 CAGCTGATCT CTCTCTCTGT GCACTCGTGA TCCATGTTGA ACAATACATG TAGGTTCTTT 1140
 TTCCACGCAA TGTAAGAACA TGATATACTG TACGTTGGAA AGCATTACG TTATTTATAT 1200
 ACCTGAATGT TCCTACTACA CAAATAAACA TATATTAAAT WCTAAAAAAA AAAAAAAAAA 1260
 45 CTGGAGGGGG GGCCCGGTAC CCAAATCGCC GGATAGTGAT CGTAAAC 1307

50

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1624 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

	GGCAGGAGGT CCGCGCGCGG GCGCGCTGGA ATTGTGAGAG TTGTGTCTGC CACTCGGCTG	60
	CCGAGGGGGA AGGTCCCTGA CTATGCTTC CCAGAGCTG CTTTCATCTA GGATCGCTCC	120
5	TCGCGGCATG CTGCTT333C TGCTGATG3C CGCTGCTTC A3CTTCTGCC TCAGTCATCA	180
	GAACCTGAAG GACTTTGCCC TGATTAACCC AGAGAAGAGC AGCATCAAAG AAAC3AGAG	240
10	AAAAGAAACC AAAGCGGAGG AGGAGCTGGA TGCGAAGTC CTGGAGGTGT TCCACCGAC	300
	GCATGAGTGG CAGGCCCTTC AGCCAGGGCA G3CT3TCCCT GCAGGATCCC ACGTACGGCT	360
	GAATCTTCAG ACTGGGAAA GAGAGGCAAA ACTCCAATAT GAGGACAAGT TCCGAAATAA	420
15	TTTGAAGGC AAAAG3CTG3 ATATCAAGAC CAACACCTAC ACATCTCAG3 ATCTCAAGAG	480
	TGCACTCGCA AAATTCAAG3 AG3333CAGA GAT3GAGAGT TCAAAGGAAG ACAAGCGAAG	540
20	GCAGGCTGAG GTAAAGG33C TCTTCGCGCC CAT3GAGGAA CT3AAGAAAG ACTTTGATGA	600
	GCTGAATGTT GTTATTGAGA CT3ACATGCA GATCAT33TA CG3CTGATCA ACAATTTCAA	660
	TAGTTCCAGC TCGAGTTT33 AAGAGAAGAT TGCTCGCTC TTTGATCTT3 AATATTATCT	720
25	CCATCAGATG GACAATG33C AGGAGCTGCT TTCCTTT3GT GGTCTTCAA3 TGGT3ATCAA	780
	TGGGCTGAAC AGACACAGAG CCTTCGTGAA GGAGTATGCT GCGTTTGTGC TGGCTGCTGC	840
30	CTTTTCCAGC AAGCCCAAG3 TCCAGGTGGA GGCCATCGAA GCGGGAGCCC TGCA3AAGCT	900
	GCTGTGATC CT3333CAG3 AGCAGCCGCT CACTCGAAAG AAGAAGCTCC TGTTTGCACT	960
	GTGCTCCCTG CT3333CACT T333CTATGC CCAGCGGCAG TTCTTGAAAC TCGG3333CT	1020
35	GCAGGTCTG AG3AGCTTG T333AGAGAA G333ACGGAG CTGCTG33CG TCGG3GTG3T	1080
	CACACTGCTC TAGGAGCTCG TCA33GAGAA GAT3TTCCGC GAGGAGGAGG CTGAGCTGAC	1140
40	CCAGGAGATG TCC33AGAGA AGCTGCAGCA GTATG33CAG GTACACCTCC TGCCAGG3CT	1200
	GTGGGAACAG GCGTGT33CG AGATCAC33C GCACCTCCTG GCGCTGCCCG AGCATGAT3C	1260
	CGGTGAGAAG GT3333CAGA CACT3333CT CCTCC33ACC A3CTG3333G ACCCTACT33	1320
45	TCAGGACCCC CAGCTCG33A G333AT333C CAGCTG33AG 33TGAGTACC AG333CT33C	1380
	CAGCTG33AG CT33333ATG GT3333333A G3333333CTC T3333333CT33 T3333333CT33	1440
50	CAACAGCTTG CT3333333C T3333333G CCGCACACCA G3333333CT GCGATGCC3C	1500
	TAGTGAGGCT GAGG33333C A3333333GT GCGTT3333C 3333333GAC ATCTTGGCAG	1560
	TGCTGCTTG G3333333AT G3333333CTA AGG3333333 AAAAA3333 AAAAA3333	1620

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2003 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

10	GGCACGAGCC AGCTTCAGG AGGAATCGGT GAGGTCTCTT CCTGAGGCTG CTGTCCGGGG	60
	COGGTGGGTG CCTCAAGGT GCTTCCCTA GCTGCTCCGG TTGCCATTGG TTCTTGCTTG	120
	TTCTGGGATG AGGAGCTTG ATTSASTTGC ACAGCTTTGC TTATCCGGG CTGTGTGTGA	180
15	GGGCCCCGGT GGGTCCCCA TGTGCACATC CTGAGGACAG AAAAAGCTGG GTCTTGCTGT	240
	GCGCTCCGAG GCTTAGTGTG CCTTCCCTCA AAGACTGACA GGCATCGTTG TGAAGGGGGC	300
20	TTCTTGATG TGAAGGAGG TAAGCATAGT AAGAAGTCCA GCTTAGGAAG GGAAGGATTT	360
	TGAGGTAGG TGGCTTTGGT GACACACTCA CTTCCTTTCTC AGCCTCCAGG ACACTATGGC	420
	CTGTTTTAAG AGACATCTTA TTTTCTAAA GGTGAATTCT CAGATGATAG GTGAACCTGA	480
25	GTTGCAGATA TACCAACTTC TGCTTGATAT TCTTAAATGA CAAAGATTAG CTAGCTAAGA	540
	AACTTCTTAS GGAAGTAGGG AACCTATGTG TTCCCTCAGT GTGCTTTCTT GAAGGCCAGTG	600
30	ATATGGGGGT TAGGATAGGA AGAAGTTTCT CGGTAATGAT AAGGAGAATC TCTGTTTTCC	660
	TCCCACTCTG GTTGTAAAGA TAACTGAGG ATATACAGGC ACATTATGTA AACATACACA	720
	CGCAATGAAA CCGAAGCTTG GCGGCTCGGG CGTGGTCTTG CAAATGCTT CCAAAGCCAC	780
35	CTTAGCTCTG TGTATTCAGG GGCAACCCCA AAGCACCTGT TAAGACTCTT GACCCCAAG	840
	TGGCATGCA GCGGATGCG CACCGGACCC TGGTCAGCAC AGATCTTGAT GACTTCCTTT	900
40	TCTAGGGCAG ACTGGGAGGG TATCCAGGAA TCGGCCCCCTG CCCCACGGGC GTTTTCATGC	960
	TGTACAGTGA CCTAAAGTTG GTAAGATGTC ATAATGGACC AGTCCATGTG ATTTAGTAT	1020
	ATACAACTCC AGGAGACCCC TCCAACCCAT ATAACACCCC ACCCCTGTTG GCTTCTGTGA	1080
45	TGGTGATATC ATATGTAACA TTACTCCTG TTCTGCTGA TTGTTTTTTT AATGTTTTGG	1140
	TTGTTTTTTG ACATGAGCTG TAATCATTC TGTGCTGTGT TTTTATTAC CCTTGGTAGG	1200
50	TATTAGACTT GCACTTTTTT AAAAAAGGT TTCTGCATCG TGGAAGCATT TGACCCAGAG	1260
	TGGAACGGCT GGCTATGCA GGTGGATTCC TTCAGGTCTT TCCTTTGGTT CTMTGAGCAT	1320
	CTTTGCTTTC ATTCTGTCTC CGTCTTTGGT TCTCCAGTTC AAATTATTGC AAAGTAAAGG	1380
55	ATCTTTGAGT AGGTTCGGTC TGAAAGGTGT GGCTTTTATA TTTGATCCAC ACAGTTGGT	1440
	CTTTTAACCG TGCTGAGCAG AAAACAAAAC AGGTTAAGAA GAGCCGGGTG GCAGCTGACA	1500
60	GAGGAAGCGC CTCAAATACC TTCACAATAA ATAGTGGCAA TATATATATA GTTTAAGAAG	1560

GCTCTCCATT TGGCATCGTT TAATTTATAT GTTATGTTCT AAGCACAGCT CTCTTCTCCT 1620
 ATTTTCATCC TGCAAGCAAC TCAAAATATT TAAAATAAAG TTTACATGCT AGTTATTTTC 1680
 5 AAATCTTTGC TTGATAAGTA TTAAGAAATA TTGGACTTGC TGCGTAAAT TAAAGCTCTG 1740
 TTGATTTTGT TTCCGTTTGG ATTTTGTGGG GAGGGGAGCA CTGTGTTTAT GCTGGAATAT 1800
 10 GAAGTCTGAG ACCTTCCGGT GCTGGAACA CACAAGAGTT GTTAAAAGTT GACAAGCAGA 1860
 CTCCGCATGT CTGTGATGCT TTGTATCATT CTTGAGCAAT CGCTCGGTCC GTGGACAATA 1920
 AACAGTATTA TCAAAGAGAA AAAAAAAAAA AAAAACTCG NCGGGGGGCC CGGTACCCAA 1980
 15 TTGCCCCTAT AGTGAGCCNA TTC 2000

20

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 786 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

GGCACAGCGG CACGAGAAGA CTTTGGTCTT TAAGAGATTA ATGTGTTAGC CAGAACAAC 60
 CATTCTCTTA CCMGTGTGTA GTCCATTAT CTTTAAAGAT TTTCTATTGG AATAATTTTG 120
 35 AAATTACTTT CTTAGTTTTC TTCATTAAAA ACTAAGAAAA TGCTTTGTTT ATTATGAATT 180
 GCTATTCTTC TTGATTATTA TTCTTGGAGA AAGTCTATCA GACGTAATTC TTCTGATTTG 240
 CTTCTAGGCT AGAGGAAAAT GTGAAAGATG ACAAATGAAA ATTTCAAAGG TTGTCAGTAG 300
 40 TATGACTTCT TTTATCGTTT GTCATTATCA CAAATATATC AACATAGGAC TTTTAAAAGA 360
 TATTTTGTAC ATATTTGGGC TTATAGGAT TTTGCATGAA TTTTMTTTT CTTTATGCC 420
 45 CAGAGAGAAA GAGCAAAGAA ATAACCAAGG GTGATGTACT GTATTGAAG GTTTACCAAA 480
 TAAGGACTGC TTTTATTATG AACTATAGTC TATATTCTAA GTAAATCAAT TTTCTATTA 540
 TGTGTTTTTT GTTCTGCAG GCAAGATCTC TGAACTTTAT GCAGAGGCTT CTTTAAAAA 600
 50 AACAAAGTTG AATTTTTTTA TTCTTGGAA TATTTMTTT CATIGATTTC TCCCAAGTAG 660
 AGCAGATTCA AATCTCTTT GTACCTATG TCTTTTGT TTGCTATTA GTCAGTATT 720

60

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1758 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

GGGACGAGCC CTGDDCACCT CCTGCAGCCT CCTGGGCCCC GCGGAGCTGG CCGATGAGC 60
 TCGCCACGGG GAGDSTGGGT AGCCAGGCGG TGGCGCGGAG GATGGATGGG GACAGCGAG 120
 15 ATGGCGGGGG CGGGAAGGAG GCCACCGGCT CGGAGGACTA CGAGAAGCTG CCGACTAGGG 180
 CTTCCGTGTT CACCCACATG ACATCAGGAG CGATGGCGGG GATCCTGGAG CACTCGGTCA 240
 20 TGTACCCGGT GGATCGGTG AAGACAGGAA TGCAGATTTT GATCCAGAT CCGAAAAGCG 300
 ACTACACAAG TATCTAGGGA GCGCTCAAGA AAATCATGCG GACCGAAGCT TCTGGAGGCG 360
 25 CTTCGAGGGT GTCAACGTCA TGATCATGGG TGCAGGCGCG GCGCATGCGA TGTATTTTGG 420
 CTGCTATGAA AACATGAAAA GGAATTTTAA TGACGTTTTT CACCAACAA GAAACAGCCA 480
 CCTAGCCAA GGTATTTTGA AAGCGTTTGT CTGGAGTTAG AAAGTTCTCT TCTTCAACAG 540
 30 GTCCCTCCCG AGGCTGTTCC TCGCTGTGAC CCAGCGCGCT CGATTCGCG CCGCTTGGTC 600
 ACGAATAAAG AACTCAGAGT TGTGTGTGCA ATGCACACCG AGACACAGCG ACGCACACAG 660
 35 AGGCGCGCGT ACACACATGC TTTTCTGTGT TCCCTCCGCG TTTGTGAAGC CTGGGAGAGAA 720
 ATCAGTGACA GAGGTGTTTT GTTTTATTG TTAATGCGGT TTTCTTTTGT ATTTTTTTTG 780
 TTTGTTTTGT TTTTAAACAT TCAAAAGCAA TTAATGATCA GACATAGGAG AAACCTGAA 840
 40 TAGAAACAAA ACTTTTGAAT GCTGATTC AAAAAAAAAA AAAGTTATCT GGACAGCTTC 900
 TTTGAGACTA TTTAAAACT GGTACAACAG GTCTCTACAA CGCCAAGATC TAACTAAGCT 960
 45 TTAAAAGGTC AAGAAGTTTT ATGCTGACA AAGGACTCGC GCAACGCAGA AGGCCTTTCC 1020
 CACCTTAAGC TTCCGGGGAT CTGGAATTT TACCCCATC CTCTCTGTT TGTCTGAGTC 1080
 TCATCTCTCT GCAAGCAAGG GCTGAAATCA TTTGTTTGG TTTTTTGGAG GGAGAGAGGC 1140
 50 GGGTGCGGG GGTGCAAAAT TCCAGCAGC TCTTACGTAA GGCATGTTTT ATTGGGAGG 1200
 GCTGAGCTTT TATTTCTCC TCTCCAGTGG GGTGGCTTT TATTGTTTCT TGTTTGGGTT 1260
 TGAATGGAA ATATGGATAG CAGCATAAAG TACTTTTATT TTGACAAAAT TCATTTTTTT 1320
 55 CAACAATGGA GACATAGATT TGACCCACAA TAACCTCTCC CCTCTCTTT TACTCTGCT 1380
 CAAAAAGCAT CTCTCTCCCT ATTACCCAAC CTTGGTCATA AGTGTGCCTG GCTGGTTTGC 1440
 60 AGATATTTGT TCTGCTTTGT AAAAAATGGC CATTAGTGCA TTTATTGAGA TGATCTCTAA 1500

AGAGCTATSC CCTGACCTAC CCTGATTCT ATGACATTGG GGGCTTTCTT TTGCTGAAAC 1560
 TGCCTTACST AATGSTMTTA CTCCTTGAAA GAGATTGAC GGAATCCATT TTATGCCAAG 1620
 5 TCTGCGCTTG CACTGTTTCT GAAATATGTS GTSTATGCTG TGCTSATCTT CCTGGGAATG 1680
 ATTATAADTS TGTGTGTGGT GGGGAGTGS GTATTACATG CATTECTGAA GAGTCAAAAA 1740
 10 AAAAAAAAAA AAACCTGA 1758

15 (2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 884 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

25 CTGTTAGAAAT GCCCAGTTTA CCTGGATGGC AACCCAACAG TGCTCCTGCC CACCTGCCCC 60
 TCAATCCTCC TAGAATTGAG CCCCCAATTG CCCAGTTACC AATAAAAACT TGTACACCAG 120
 CCCCAGGAGC AGTCTCAAAT GCAATCCAC AGAGTGASMC ACCACCTCGG GTAGAATTG 180
 30 ATGACAACAA TCCTTTTAGT GAAAGTTTTC AAGAACGGGA ACGTAAGGAA CGTTTACGAG 240
 AACAGCAAGA GAGACAAAGG ATCCAACTCA TCCAGGAGGT AGATAGACAA AGAGCTTTGC 300
 35 AGCAGAGGAT GGAAATGGAG CAGCATGGTA TGGTGGGCTC TGAGATAAGT ASTAGTAGGA 360
 CATCTGTGTC CCAGATTCCC TTCTACAGTT CCGACTTACC TTGTGATTTT ATGCAACCTC 420
 TAGGACCTCT TCAGCAGTCT CCACAACACC AACAGCAAAT GGGGCAGGTT TTACAGCAGC 480
 40 AGAATATACA ACAAGGATCA ATTAATTGAC CCTCCACCCA AACTTTGATG CAGACTAATG 540
 AGCGAGGCAG GTAGGCCCCC CTCATTTTGT TCCTGATTCA CCATCAATCT CTGTTGGAAG 600
 45 CCGAAATTTT TCTTCTGTGA AGCAGGGATA TGGAAATCTT TGTGAGACA CTTTCAGCA 660
 GTCCCCAGTG AGGCTTTCTT TTACACCTGC TTATCCAGCA GCACCTCCAG TAGCTAATAG 720
 CAGTCTCCCA TGTGGCCAAG ATTCTACTAT AACCCATGGA CACAGTTATC CCGGATCAAC 780
 50 CCAATCGCTC ATTCAGTTGT ATTCTGATAT AATCCAGAG GAAAAAGGN AAAAAAARA 840
 AMAAFAARA ARAAAGGAGA TGATGATCCA GAATCCACC AAGGCTCC 888

60

(i) SEQUENCE CHARACTERISTICS:

426

(A) LENGTH: 2379 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

	GGCAGAGCTA GTGTGGACTC CATCCCTCTG GAGTGGGATC ACGNCTATGA CCTCACTCGG	60
10	GACCTGGAGT CTGCAATGTC CAGAGTCTG CCTCTTCAAG ATGAAGAAGG TCAGGATGAC	120
	AAAGATTTCCT ACCTCCGGGG AGCTGTGTTT TTATCAGGAG ACCACAGTGC CCTAGAGTCA	180
15	CAGATCCGAC AATTGGGCAA AACTTGGATG ATAGCGGCTT TCAGATACAG TAAACCGAAA	240
	ATATCATTTG CAGCAAAACT CCGACGGGGC CCGAGCTAGA CACCAGCTAC AAAGGCTACA	300
	TGAAACTTCT GGGGCAATGC ACTAGCAGTA TAGACTCTCT GAAGAGACTG GAGCACAAC	360
20	TGAAGGATGA AGAGGAGAGC CTTCCTTCTT TTCTTAAGCT GATAGTAGGC TAAACCGAAA	420
	CGGCTGCTCT CATTCACCGA TCGGAGTTTC TCGAGGCTCA GGCATTGAGC AAGGAGTTGA	480
25	GGATGAATCA GAACCTCCAG AAGTGGTAGC AGTTTAAGTC AGACTTGAAC AGCATCTGGG	540
	CCTGGCTGGG GGCACAGGAG GAGGAGTTGG AACAGCTTCA GGTCTCTGGAA CTCAGCACTG	600
	ACATCCAGAG CATTCAGCTC CAGATCAAAA AGCTCAAGGA GCTCCAGAAA GCTGTGACCC	660
30	ACCGCAAAGC CATCATCTCT TCCATCAATC TCTGCAGGTC TSAGTTCAAC CAGGCTGACA	720
	GCAAGGAGAG CCGGCACTG CAGGATCCCT TCTGCAAGAT GAATGGGCTT TGGGACCGAG	780
35	TCTGCTCTCT GCTGAGGAG TGGGCGGCTT TCTGCAAGGA TCGCTGATG CAGTGGCAGG	840
	GTTTCATGTA AATGAGCCAT GCTTTCTCTT TTATGCTGTA GAACATTGAC AGAAGGAAAA	900
	ATGAAATTCT CCTATTGAT TCTAAGCTTG ATGCAGAGAT ACTTCAGGAT CATCACAAC	960
40	AGCTTATGCA AATAAAGCAT GAGCTCTTGG AATCCCAACT CAGAGTAGGC TCTTTGCAAG	1020
	ACATGCTCTG CCAACTACTG GTGAATGCTG AAGGAACAGA CTGTTTAGAA GCGAAAGAAA	1080
45	AAGTCCATCT TATTGGAAAT CCGCTCAAAAC TTCTCTTGAA GGAGGTCAAT GGTCAATCA	1140
	AGGAAGTGTA GAAGTTATTA GAGCTGTCAA GTAGTCAGCA GSAATTTGTCT TCCTGGTCTT	1200
	CTGCTGATGA ACTGGACACC TCAGGCTCTG TGAGTCCCAAT ATCAGGAAGG AGCACCCTCA	1260
50	ACAGACAGAA AACGCCACGA GGCAGTGTGA GTCTCTCAGA GCTGGACCTT TCTGTGAGCA	1320
	GTCCACATAG CAGGTCCACA AAAGGTGGCT CGGATTCTCT CTTTCTGAG CCARGGCCAG	1380
55	GTGGTCCGG CCGGGCTTTC CTGTTGAGAG TCCTCCGAGC AGCTCTTCCC CTCAGCTTTC	1440
	TCCTGCTCTT CTTTATCGGG CTTCCTGCTT TTGTACCAAT GTCAGAGGAA GACTACAGCT	1500
	GTGCTCTCTC CAACAACCTT GCGCGTCTAT TCCACCCCAT GTCAGATAC ACGAATGGCC	1560
60	CTCTCCACT CTGAACCTAAG CAGATGCCAT CTGCAGAAGT GCTGGTAGCA TAAGGAGGAT	1620

5 CCGGTCATAA GCAATCCCAA ACTACCAACA AGAGGACCTT GATCTTGGCG AAAGCCMTCC 1680
 GTGTGGCAGC TTTAGCCTCC TCAGATCAC ATGTGTCCAA ATTATGGCTT CAGAGGTGGA 1740
 AGATAAACAG TGACGGGGGA ACAAACAGAC AACAGAAGG TTTGGAAGAA ATCTGGTTTG 1800
 AGACTCTGAA CCTTAGCACT AAGGASATTG AGTAAGGACC TCCAAAGTTC CCGGACTCA 1860
 10 TGAATTCTGG GCCCTTGGCC NATTCGTGTC ACAGCCAAGS ACTTCAGTAG ACCATCTGGG 1920
 CAGCTTTCCC ATGGTGTGTC TCCAACCATC AGATAAATGA CCTTCCAAG CACCATGTCA 1980
 GTGTCTACA ATCTACCAAC CAACCACTGC TGAASAGATT TTAGAACCCT GTAACATACA 2040
 15 ATTTTAAAGA GCTTATATGG CAGCTTCTTT TTTACCTTGT TTTCTTTTCG GGCATGATGT 2100
 TTTAACCTTT GCTTAGAAG CACAAGCTGT AAATCTAAAA GGCATTTTTT TTAGAGGTA 2160
 20 TAAAGAAAAA CTAGATGTAA TAAATAAGAT CATGSAAGGC TTTATGTGAA AAAAGTTGAA 2220
 TGTATAGTA AAAAAAAG ATATTTATST ATGTACAGTT TGCTAAAGCC AAGTTTGT 2280
 TGTATTGATT TCTTTGCAT TATTATAGAT ATTATAAAT AAAAAAAAAA AAAAAAAC 2340
 25 TCGACGGGGG GCCCGTACC CAATTCGCC TATAGTGAG 2379

30

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 1343 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

40

CCGGCTTCAC GATCCGGGCG GTCAGTGGTC CAGGTCCCTT ATTCTGCCTT CTCCTCTGCC 60
 TCTTGAATCC CCATACCTT GAGACGGGT GTCTCTCTGT ACCGAGTTT GATATAAGC 120
 45 TCAGCTTCAA AGGCTCAAG CTGGCATTC CTGAGCTGG AATACCTTC TGAAGCATC 180
 ATGSAAGTGA GGGCAGGGG TGGGAGGCG TATGCCAGG GTCCCTCAAA TGCTGGAGG 240
 GGCTGTRACT TGTGCGGAG TGGGTCTGTC ACAGCATCC TCTGTCCAGG GTGGGGCAAG 300
 50 GCTTGGGACA GTCCAGGCA CCCCAGGACC CTTCCAGGC TTGTCTCTG CTCACCGCC 360
 TTAACAGCCG CAGCTCTTC CAGCTCTTC TCTCTCTGC CTCTCTCTTC CCGAGGCA 420

60

TAAAGAAAAA GTCAGTTT GAGCTTCA TATAGTGG GATCTCTTC TATATAAAAA 480

	CTGGGGGCTA CCTGGAGGGA AGCATCTCA TCCAGGTGA GTGGGACCA GCGCTTCCT	660
	GTATGTGTGT TGTGGGTGGA AGTAGGCATG AGAGCATCTT AGCCATAGG TTTGTATTCA	720
5	GGGACTTCCA AACCCAGACC TAJAAAGAGT GTGTCTTCTA CAGATCTTG TTCAAAAAG	780
	GGTTTGTGAT GATGSAACTA CACCATAGAG GASTGAGCA AGAAATGA GGATTAGAT	840
10	GGAGCGTGA AATGTATAGG AGCATGGCTT CAAAACATA TGTGTGAGG TCTGTCACG	900
	AGAGATTGAG GGCATGATT TATTTCTGAG CTTCTTAGCA GCAAAAGCA AGACAGAAAG	960
	CAGATGGCT GTGGATTCTT GTTATAAAA TGTAGTTCT TGGCGGCTT CGGTGGCTCA	1020
15	GGGTGTAAAT CCGGCTGCTT TGGAGGCGCA GGGGGGAGG GTGGAGAGT CAGGAGGTTG	1080
	GAAACCATCC TGGCGGAAT GGTGAAGGCT TGAATCTACT AGAAGTCAA ASATTGGCTG	1140
20	GGTGTGTTGG CGTGGGCTG TGTGCGAGG TTCTGGGAG GTGAGGCGG GAGATTGCT	1200
	TGGCGGAGG AGGCGAGGT TGGGAGAGT TGATGCTG CATTGCACT TCAGTCTGG	1260
	GACAGAGCGA GACTCTGGCT CAAAAA AAAA AAAA ACTGAGGG GCGCGTACC	1320
25	CAATTGCGG NATATGATG TAAACAAT	1348

30 (2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1502 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

40	CTCAAAATAA ATAAATAAAT AAAAATTTGT ATTCCATTGA TTTGGGTAGA CACCAAGAAT	60
	GTGCATTTCT AACAGCTTT CCAGCGATC CTATAGTAAG TCATCTGTG ACTACTTTAA	120
45	GAAACTCTTC TATAGAGAAT GGAGTTGGAT TAATAATAGG TGATTTTFTA CACTGACTG	180
	ATTCACAAGA ACCTAAACAG TAGTGCATGA AGCTGCTCAT CTGTGTAAG TATTTGCCC	240
	CGTCTCACTC TGAAGCAGC AGGAGATGT GTTACTTTG TTCTATGCG GTTGTCTGG	300
50	AGATTAATTT TGAATGAAA GTTTTCTCT CTATGCATT CCTGGTTCTT TTCAAAGCC	360
	TCATACAAGA GGATTAGGTC AATATGATG CATTACTTT TAAAGAATG CGATATTGAT	420
55	ACCGATGCTT ACTTTTTTTT TTTTINACTA CTGTGTTTAT TCCTTCAGN AAAGTATAGC	480
	CCGCCCTTCT ATAGCATAGT TCTCTTAGG TGAATGATT CCTATAAGAT TTCTATTAT	540
	TAAATCATGC ATTTTCAAG ATGAATCAA TMTTGATTT AATCTAAGCT GATATTCTCA	600
60	TTTGTTAGAA GAACAACCTA CATGCTAGG AGAGAGGAGG AAATATACCC ACGACCACAC	660

AGCCAGTTAG TATCCAGTTG GTGCTGGACT CCAGCCAGGT GTGCTGCCCTC ATGGTAGTTA 720
 AATGATATAT AGAAAAGGTA AATTTTAAA GAAATAATTA TTAATATATT CCTATAAAGC 780
 5 ATTTTAAAGG TAACCACATA AAATGGTTA ATTTTCCAT TCCAAAGTAA ATGCTAAGCA 840
 TGTTTATTAA TGAAGCAGTA CTTCTGATTA GTATATGACA TTCTGAASIT AATTAAACTC 900
 10 ATTCAGCTAA ATGTSTCTTC CTTGTATAG TGGAGGATTT GAGGATTTGA ATATAGAGTA 960
 GAGTGCTTGC TTAAGCCTGG GAGGCCATCT TTATAGCTAT TTGATGTAAG AAAAGAGACA 1020
 TGSNCCATTT CTAACTATA TAACTGAGT GTCTCTATTC CCAGCAGATA TAAAGGAAAA 1080
 15 AGGAAACTTT TTTGATTCCT ACCTTCCAG CCTCAGCTAG CCATCTTCCA CCTCAAATA 1140
 TAGAGATGTT AGTGCAAGGT CCTGGCTCT AGGTGATCAT TTCATAAGTC CTTTACAGAT 1200
 20 AAAGAAAAAG TAGTGTGTGT ATTTTGTTT TTAAGTAACC CCAAAACAAA TTTATATTGT 1260
 ATTCAGCAAA ATTGSAATTC AGGTGTTTAA TTTTGAACA TGAAGTCCCT CCGTGTTTTAA 1320
 GCATTGACTT GTATAAAAAA AATTGCATGT CTCCAGTAAG CTTATGGGTT TTCTCATTTT 1380
 25 TAGGTATATG GCTTTTAATC ATGTAAAGTG AACATTAGT TTTCTTGCAT TTTATTACAG 1440
 GTTCTTTGTT GCAATAAAGA TGCTCTGAA ATTAATTGAA AAAAAAAAAA AAAAAAATC 1500
 30 GA 1502

35 (2) INFORMATION FOR SEQ ID NO: 178

(i) SEQUENCE CHARACTERISTICS.

(A) LENGTH: 1637 base pairs
 (B) TYPE: nucleic acid
 40 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

45 ATTTCTTAGC CCADAAGGAC TGAAGTTGAG ATCCAAAAGT TCACTTGATA ATTATCTTCA 60
 CAAAAATGSA GAGACTTCTC TTAAGCCACA AATTTTGTAT TTTACTGTAC TTTCTAAAG 120
 GGGTATCAAG TCAAGATATA AAGACTGAG CATGGCAGCC CTGACATGCC ATCTACAAAA 180
 50 CCAAAGTAAC AATTCAAAC TGAAGTTGAG GACCCGAGC AAGTGCAAAA AGGATGTGTT 240
 TATGCCGCCA AGTAGTATT CAGAGTTGCA GGAGAGCAGA GACTCTCTA ACTTTACTTC 300

60 ATTAAGGAG AATTTTATG GTTCTTTT AATGATAA AAAAGATAAT TTTCTTAA 480

	TAAAGCAGAT GCTGAAAGTG AACCTGTTGC ACAAAAAAAT CAGCTTGATA GAAGTGTCTG	540
	CATTTCTGAT GCTGGAGGAT GTGGTGAGAC CCTCAGTGTG ACCAGTGAAG AAAACAGCCT	600
5	TGTAAAAAA AAAGAAAGAT CATTGAGTTC AGGATCAAAT TTTTGTCTTG AACAAAAAAC	660
	TTCTGGCATT ATAAACAAAT TTTGTTGAGC CAAAGACTCA GAACACAACG AGAAGTATGA	720
10	GGATACCTTT TTAGAATCTG AAGAAATCGG AACAAAAAGT GAAGTTGTGG AAAGGAAAGA	780
	ACATTTGCAAT ACTGACNTTT TAAAACTGG CTCTGAAATG GACAACAATC GCTCAGCAAC	840
	CAGGAAAGAC TTCACTGAAG ATACCATCCC AGGGAACACA GATAGAAAGA AGGAAAAACA	900
15	GCCTGTATTT TTCCAGCAAA TATAACAAAG AAGCTCTTAG CCCCCCAGCA GGTAAAGCCT	960
	TTAAGAAAGG GACAGCTCTT CGGTACCTTT TTAATCTGCT TCAAGAAACA CTTTTTCATG	1020
20	ATCCATGAAA GTTTCTCATG GGTACTATAT TTTCGAATGG GACCTCAGGC AAAATGGGAA	1080
	TACCTGTGCT TTGGAAGTTT TGGGAGAAGT ATCTTCAAC TGAGGTAGGA AGAACCOCAG	1140
	ACTGGAGAGA TTTCTAGAA CTTCTTAAAC CTCTTGGTCT CTACGATCTT CCGGCAAAAA	1200
25	CCATTGTGAA GTTTCTCAGT GAATACCTGA CAAAGCAGTG GAAATATCCA ATTGAGCTTC	1260
	ATGGGATTGG TGAACCTTGA AGACACAAA TTAAATAAAT ATCATGACTG GCTTTGGGAA	1320
30	AATCATGAAA AATTAGTCT ATCTTAACT CTGCAGCTTT CAAGCTCATC TGTATGCAT	1380
	AGCTTTGCAC TTCAAAAAAG CTTAATTAAG TACAACCAAC CACCTTTCCA GCCATAGAGA	1440
	TTTTAATTAG CCAACTAGA AGCCTAGTGT GTGTGCTTTC TTAATGTGTG TGCCAATGGT	1500
35	CGATCTTTGC TACTGAATGT GTTGAACAT GTTTTGAGAT TTTTTTAAAA TAAATTATTA	1560
	TTTGACAACA ATCCAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1620
40	AAAAAAAAA AAAAAAA	1637

(2) INFORMATION FOR SEQ ID NO: 179:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2911 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

50

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

55	GGTGGTTTTT GTTCTGCAAT AGGCGGCTTA GAGGGAGGGG CTTTTTCGCC TATACCTACT	60
	GTAGCTTCTC CACGTATGGA CCCTAAAGGC TACTGCTGCT ACTACGGGGC TAGACAGTTA	120
	CTGTCTCAGC TCTAGGATGT GCGTTCTTCC ACTAGAAGCT CTTCTGAGG3 AGSTAATTAA	180
60	AAAACAGTGG AATGGAAAAA CAGTGCTGTA GTCATCCTGT AATATGCTCC TTGTCAACAA	240

	TGTATACATT CCTGCTAGGT GCCATATTCA TTGCTTTAAG CTCAAGTCGG ATCTTACTAG	300
	TGAAGTATTC TGCCAATGAA GAAAACAAGT ATGATTATCT TCJAACTACT GTGAATGTGT	350
5	GCTCAGAACT GGTGAAGCTA GTTTCTGTGTG TGCTTGTGTC ATCTCTGTGT ATAAAGAAAG	420
	ATCATCAAAG TAGAAATTTG AAATATGCTT CCTGGAAGGA ATTCTGTGAT TTCATGAAGT	480
10	GCTCCATTCC TGCTTTCTTT TATTTCTGTG ATAACCTTGAT TGTCTTCTAT GTCTGTCTCT	540
	ATCTTCAACC AGCTATGGCT GTTATCTTCT CAAATTTTAG CATTATAACA ACACTCTCTC	600
	TATTCAGGAT AGTCTGAAG AAGCGTCTAA ACTGATCCA GTGGGCTTCC CTCTGACTT	660
15	TATTTTGTG TATTGTGGC TTGACTGGCG GGAATAAAC TTTACAGCAC AACTTGCAG	720
	GACGTGGATT TCATCAGAT GCTTTTTCOA GCTTTTCCAA TTCCTGCCTT CTTTTCAGAA	780
20	ATGAGTGTCC CAGAAAAGAC AATTGTACAG CAAAGGAATG GACTTTTCCT GAAGCTAAAT	840
	GGAACTAC ACCTCAGAGT TTCACTCACA TCCGTCTTGG CATGGGCCAT GTTCTTATTA	900
	TAGTCCAGTG TTTTATTTCT TCAATGGCTA ATATCTATAA TGAAAAGATA CTGAAGGAAG	960
25	GAAACCACT CACTGAAGC ATCTTCATAC AGAACAGCAA ACTCTATTTT TTTGGCATTC	1020
	TCTTTAATGG GCTGACTCTG GGCCTTCAGA GGAGTAACCG TGATCAGATT AAGAACTGTG	1080
30	GATTTTTTTA TGCCACAGT GCATTTTCAG TAGCCCTTAT TTTTSTAAT GCATTCAGG	1140
	GCTTTTCAGT GCTTTTCATT CTGAAGTTC TGGATAACAT GTTCCATGTC TTGATGGTCC	1200
	AGGTACCAC TGTATTATC ACAACAGTGT CTGTCTGGT CTTTGACTTC AGGCTCTCC	1260
35	TGGAATTTTT CTGGAAGCC CCATCAGTCC TTCTCTCTAT ATTTATTTAT AATGCCAGCA	1320
	AGCCTAAGT TCGGAATAC GCACCTAGGC AAGAAAGGAT CCGAGATCTA AGTGGCAATC	1380
40	TTTGGGAGCG TTCCAGTGGG SATCGAGAAG AACTAGAAAG ACTTACCAAA CCAAGAGTG	1440
	ATGAGTCAGA TGAAGATACT TTCTAATGG TACCCACATA GTTTGCAGCT CTCTTGAACC	1500
	CTATTTTCAG ATTTTCAGTG TTCTAATAT TTATCTTTTC ACTTTGATAA ACCAGAAATG	1560
45	CTTCTAAATC CTAATATTCT TTGCATATAT CTAGCTACTC CCTAATNGGT TCCATCCAAG	1620
	GCTTAGAGTA CCAAGAGGT AAGAAATTC AAAGAACTGA TACAGGAGTA ACAATATGAA	1680
50	GAATTCATTA ATATCTCAGT ACTTGATAAA TCAGAAAGTT ATATGTGCAG ATTATTTTCC	1740
	TTGCTCTTCA AGCTTCCAAA AACTTGTA TATCATGTT AGCTATAGCT TGTATATACA	1800
	TTATTTTCA ATAAAGATTA AATCTCTTAAAGT ATCTTCTCTCT	1860
60	TACTTGCCA TACATAGAT TTGGAATGAT GATCTGTGTG CTAAATATTT TCTTAAGAA	1920

	GCAGTTTCTC AGACACAACA TCTCAGAATT TTAATTTTTA GAAATTCATG GSAAATTGGA	2100
5	TTTTTSTAAT AATCTTTTGA TTTTTTAAAC ATTGGTTCCC TAGTCACCAT ASTTACCACT	2150
	TGTATTTTAA GTCAATTAAA CAAGCCACGG TGGGGCTTTT TTCTCTTCAG TTGAGGAGA	2220
	AAAACTCTGA TGTCACTACT CTTGAATTAT TACATTTTGG AGAATAAGAG GGCATTTTAT	2280
10	TITATTAGTT ACTAATTCAA GTTGTACTA TTGTATATCT TTCCAAGAGT TGAATGCTG	2340
	GGTTCAGAAT CATACAGAT TGTCASTGAA GCTGATGCCT AGGAACTTTT AAAGGATCC	2400
15	TTTCAAAAGG ATCAATTAGC AAACACATGT TGACTTTTAA CTGATGTATG AATATTAATA	2460
	CTCTAAAAAT AGAAAGACCA GTAATATATA AGTCACTTTA CAGTGCTACT TCACACTTAA	2520
	AAGTCATGG TATTTTCAT GGTATTTTGC ATGCAGCCAG TTAACCTCTG TAGATAGAGA	2580
20	AGTCAGGTGA TAGATGATAT TAAAAATTAG CAAACAAAAG TGACTTGCTC AGGGTCATGC	2640
	AGCTGGTGA TCATAGAAGA GTGGGCTTTA ACTGGCAGG CTGTATGTTT ACAGACTACC	2700
25	ATACTGTAAA TATGAGCTTT ATGGTGTCAT TCTCAGAAAC TTATACATTT CTGCTCTCCT	2760
	TTCTCCTAAG TTTCATGCAG ATGAATATAA GGTAATATAC TATTATATAA TTCAATTGTG	2820
	ATATCCACAA TAATATGACT GGCAAGAATT GGTGGAATTT TGTAATTAAA ATAATTATTA	2880
30	AACTAAAAA AAAAAAAAAA AAAAAGCTGA G	2911

35 (2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

45	GGCAGGAGGC CCAGGCCAGC CAGGGCCAGG CCTACTTTGG CCACCCCTAA ATTAGAATGT	60
	GGGGTCAGGG GTACAGAAAA AGGCATTTCT CTGACCTAST GTTTGGCGTC CGGGAACCTT	120
50	GTGCCCAAGC TTCAGACCTT GGCAGTCTCT ACTGAGGCCA TTGGCCCAAG GCCCGCCATC	180
	CCCCGAGAGC CCCCCGAGGC GCCTGTTCGC ACGTCCACAC CTGCCACACC CTCTGCCGGG	240
	CCCCAGCCCT TCCCAACCGG GATGGTCTGT GTCCCTGGGG GTCCCTGCCCC ACCTTGCCCTT	300
55	GGGGAGGCAT GGGGCTCTCT CTTCCACCCC TGCCCGCCCT CACTCACTCT TTGCTTCTGG	360
	TCCCCAAGGC CTAGCCCTTG GAAGGAGACA GGAGTCTAGG GAGGCTGAAG CCCACTCCCC	420
60	GGGAGGCCCG TCTCTCTCCA GCCCCAGGGA CAGCAAGGAA AAGAGAAGAG AGCAGAGCAT	480

TTCATGCGCTC TAATAAAAAA AAAAAAAAAA AAAACTCGA

519

5

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 958 base pairs
(E) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

15

TCCCTTTGGG GCCGAAAAA GCGGGGTTGG CTTGNCATT GTTNTTCAT GCGGCCCCGG 60

CATGCCCCAG TACTAGCGTG CAGTCCCAAT GTAGCCCCCT CTCYTCCMA GAGCCCTCM 120

20

AACGCCCCCG STCAATTSTG ATTTCAGGAG GATTTCATGA AGATGTTAAA GCGAAAGTGG 180

AGAACTTTCT CCGATTTCC AGCGTGAAA AAAGCGAGCC TGTAGGCAA GCACCTGCA 240

25

GCGCTCCCTG TCCCTTTCTT CCGCTCCCTT TCYCCGCGC GTGGAGACAG CTGTTTTCAG 300

CAGGGCTCTT CCGAGGAGG GCGCCGCGCT CTTCCTTGC AGCAACATCC TTGCCCTTGT 360

CACACAAGTC AATCTCCATC TCGCCAGCTC TGTGATGCG CTGCTCGAGG GCAACAGGTA 420

30

TGTCACTGGC TGTTCAGCC CCGACACCG CCGCGGAAG CTCATCCACC CGCTCATGCT 480

TCAGCACATC CAGCCCGCAG CCGTCAGCT CCGGACAG TGGAGCACCC TCGTCAGGA 540

35

GCTGAGGCT GCCCTGCAGC TGGCTTTCTA CCGGATGCC GTGGAGGAGT GGCTGAGGA 600

AAAGCTGCAC CCGAGCTGC AGCGGCTGCA ARCTCTCTG CAGGACCTCA GCGAGGTGTC 660

TGCCCCCGCG CTCCACCCA CCAGCCCTGG CAGGACCTT GCTCAGGACC CCGAGGGGA 720

40

GAGCTCATGC CAGCGGCGTC CTCTGGAGG CTGGGGGGGC TCTGCWYTKY CWWWPGCCT 780

CGGCAATACG GCGCAGCTGG GCGTGGTGGC CTCTGCCCCA GCAGTGTCTT GCGCACATC 840

AGTTCTGAG AGGCTCGGG AGGCTCGGG GAGAGACTA GAAACACAG AAGGAAGCAG 900

45

CAGAGGGAGA CCGCTTTTGT GATCTGCATG TGTACACTG ATTCTTTGGA AATAAAGAGT 960

GGAAGCTG 958

50

(2) INFORMATION FOR SEQ ID NO: 182:

STRANDEDNESS: double
TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

TGTAAAAGTT ATCAGTAATC CTAATTCTTT TCGTGAGTTT TCGTTTGTG ACTTATTAAAT 60
 5 CAGTMTTGA AAGGACGAAT GAATTTAGAG ATGTACTCTG GAGCAGTATC ATGTTAAACC 120
 AGGGGTATAT TAGAAAAATC ATCTCATAA TCATTCTGGG AAGTTTTC TCCCAAAAA 180
 AAGCCATCCT GATGGGTTTT CAAAACCAGA AAAAGCTCT TAATGAGGAA CAGACCACTG 240
 10 GASTACCTAT GAGCATCTCA GGAAACTGA GACCTCGAG AAGCCTTGAT TCGTGCAAC 300
 CCGCAAGGTT TCAGAGGAG CAGCCCACTG CTGTCTTGA CAGACGTGGT TTTTGGGGA 360
 15 AAGCAGGAG AGGCCAGGAA TTTTCAGAGT CGTGAGTAC GRTTCCAC CCAAGATTAG 420
 AGCAGAGAT AGCATACTG AGATTCTGA AAATCATCT GTCTAAGCAA TGGAGGTGT 480
 TCGAACCTG CAGTGGCTCT TCACAGGGGA TCGAGGAGA TSYGGGTTT AGGATGGGG 540
 20 AGGCCAGGAG AGCCCTCTG AYTCTCTGC ACCTGCTCC TCAGGTGAC ACTGTCCACA 600
 ACTGTGGCTC TCACAGGAA GTTCCCAAG GAGCTCATAT CTTATTGAG ATAGGGGGTC 660
 25 GTACAGGTGA CATTCATGAG CAGTGTGAG CGCGTGACAT GGGGTGTCA ACCAGCATC 720
 TGTCCAGGAG CTCTCTCTGC AGCGGCTCT GCAGGTGGCC TGAGGCTCCT TTTTGAGAGA 780
 GAACTGTTG GCTTCTCTGT CTCTCTCTCT CTGATCTGTT CTTCTTGGG ACAACACCCA 840
 30 AGAACCTCAG CTCTCTCTAT AGATTGTGAG CTCTGGAGG GCAGGAGCTG TGTCTTCTA 900
 TTCATCTTCC TATCCCTAGA ACCTTGACA GATCTGGAA TGTGGTAGGT GCTTAGTAA 960
 35 TGTGTGTGA ATAAATGAAT GAATGAATGA ACAAATGAAT GAATTTGCTT ACTTCAAGCC 1020
 AAAAGAACCA TGAAGCTGA TTTGAGTTT CTATGTTATA CCAGTCAGCA AATCTATTA 1080
 40 AATACTTTGT GTTCCAAGC AAAAAAAAAA AAAAAAAAAA AACTCGA 1128

(2) INFORMATION FOR SEQ ID NO: 183:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2276 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

CCGCGGCGTC TGACCTCATG GCGTAGAGCC TAGCAACACC GCAGGCTCCC AGCCGAGTCC 60
 55 GTTATGGGCG CTGCCGTCCC GAAGAGGATG AGGGGGCCAG CACAAGCGAA ACTGCTGCCC 120
 GGGTGGGCA TCCAAGCCTT TGTGGGGTTG GCGGGGCGC TGGTCTTGGC GCTCTGCTT 180
 60 GTGTGGGCG CTCTATCCAG TGTGTATCA CGGACTGATT CAGGAGGCC AACCTACTC 240

	AACTCACATA TTTCTACCCC AAATGTGAAT GCTTTAACAC ATGAAAACCA AACCAAACCT	300
	TCTATTTTCC AAATCAGCAC CACCCTCCCT CCGACGACGA GTACCAAGAA AAGTGGAGGA	360
5	GCATCTGIGG TCCCTCATCC CTGGCCTACT CTTCTCTCTC AAGAGGAAGC TGATAAGAAT	420
	GAAGATCCTA GTATAGAGGA GGAGGATCTT CTCATCTCTG ACASTTCTCC ATCCACAGCC	480
10	AAAGACACTC TAGACAATGG CGATTATGSA GAACCAGACT ATGACTGAC CACGGGCCCC	540
	AGGAGCAGC ACGAGTCTGA TGACACCTTG GAAGAAAACA GGGGTTACAT GGAAATTGAA	600
	CAGTCAGTGA AATCTTTTAA GATGCCATCC TCAAATATAG AAGAGGAAGA CAGCCATTTC	660
15	TTTTTTTCATC TTATTATTTT TGCTTTTTCG ATTGCTGTTC TTTACATTAC ATATCACAAC	720
	AAAAGGAAGA TTTTCTTCT GGTTCAAAGC AGGAAATGGC GTGATGGCCT TTGTTCCAAA	780
20	ACAGTGAAT ACCATCGCCT AGATCAGAAT GTTAATGAGG CAATGCCCTC TTTGAAGATT	840
	ACCAATGATT ATATTTTTTA AAGCACTCTG ATTTGAATTT GCTTATGTAA TTTTATTTGC	900
	TTGACTTTTT ATATGATATT GTCCAAATGT TTGCCATAGC CAATTGCTAC TTAAATGAGA	960
25	GGTGACTCTC TCTTTTCCCT TGGTGCTTTG GAAATTAAT GTCACAAACG AGTATATAAT	1020
	TTTTTATCTG TACTTTTAGA GGTGAGTTTA ATCAGGTGTC CAAATGTGA GTTAAACATT	1080
30	ACCTTATATT TACACTGTTA GTTTTTATTG TTTTAGATTT ATTATGCTTC TTCTGGAAGT	1140
	ATTAGTGATG CTACTTTTAA AAGATCCCAA ACTTGTAAGT AAATCTCTAC ATATCTGTTA	1200
	CTGCTGACTC ACATTCATTC TCCGCCATTC AAATACTATT TTTTATCCAC ATTTTTTTCT	1260
35	GTTCCCAAAC TGAATGTAC AAGGATATGT GTGATAATGC TTTGSAITTG AGTAATATTT	1320
	TTTTTCTTC CAAGAAACT CCTTTGATA TTTTAGATA ATTTAAACAT AATTAGGAT	1380
40	AATGATATTG CTCAATCTGA CCACAATTTT AGGTAAAACA TTAAATGTGT CAGAAATCTT	1440
	GGCAGACAGG ACTCTGACG TTCCAGTGA CATAGATGAA ATGTTACAGA GATACTATTT	1500
	TTTCTGTTGG AATTACTATA TTAAATTTAG AAGCAGAAAC TGGTAAATG TTAAATACAT	1560
45	GTACAATTC TTTTAGTTAG CAATTCATTC TACCATGCT TCTCCAAGG TTCTAAGCAA	1620
	TGGGCAGAST TTAAATTTAT ATCAGATTGG TTTACTTCTT TTATTATTTT ACASTAAATT	1680
50	TGAATAAATC TTACGGGTCA TTATCACTTA AATAATACTG TACCTAGGTC TTTCAAATTA	1740
	AAATTATAGC TGAATGAAT TGTGTGTATA CATAAAGGAT ATTGTGTAC AATTACCTTT	1800
	TTTAAATTA TTTTACTTA TAAATTTTAA TAAATTTT TAAATTTT AATTATTA	1860
60	TATTTATCTT AATACCTGCG TGAAGGCTG TCTTCTCTT TTATTTAGAG TTTTAAAG	1920

5 GCCGTCCATC CTGTCTCTTG GGGGACAGT GTACTTTCTT AATAGGGAAG GGAAGCACAA 2100
 TGGAAATACC CCTGAACCGT TTTATTGCAG TAATTTTMTT CATATCTGAA ACTATTATTT 2160
 AATA'TTTTGA ATAAGATTTT AAAAAATAAA TGGCAAAGAT ATAAATCTAA AAAAAAAAAA 2220
 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA 2276
 10

(2) INFORMATION FOR SEQ ID NO: 184:

15 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 20
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

TCCAAGCTAC GGCACCTGGG CTGGGGCGTT GGGAGCGGGA GTGCAGAGCG TGGTCGTGGC 50
 25 GGGGGCGSTG ASAAGAGCGA GCGKAGGAG GGGGTGCCAT GGCCGGGCAG CAGTTCCAGT 120
 ACGATGACAG TGGGAACACC TTCTTCTACT TCTTCACCTC CTTCGTGGGG CTCATCGTGA 180
 TCCCGCGCAG AFACTACCTC TGGCCCCGAG ATCAGAATGC CGAGCAAATT CGATTAAAGA 240
 30 ATATCAGAAA AGTATATGGA AGGTGTATGT GGTACGTTTA CGTTTATTAA AACCCAGCC 300
 AAATATTATT CCTACAGTAA AGAAAATAGT TCTGCTTCCA GATCGGCAT TGTCTTATT 360
 35 CCTTCATAT AAAGTTTCCA AACAGACCG AGAATACDAA GAATACAATC CTTATGAAGT 420
 ATTAAATTIG GATCCTGGAG CCAQAGTAGC AGAAATTAAA AAACAATATC GTTTGCTGTC 480
 ACTTAAATAT CATCCAGATA AAGGAGGTGA TGAGGTTATG TTCATGAGGA TAGCAAAAGC 540
 40 TTATGCTGCT TTAACGGATG AAGAGTCCCG GAAAAATTGG GAAGAATTTG GAAATCCAGA 600
 TGGGCCTCAA GCCACAAGCT TTGGAATTGC CCTGCCAGCT TGGATAGTTG ACCAGAAAAA 660
 45 TTCAATTCTG GTTTTACTTG TATATGGATT GGCATTTATG GTTATCCTTC CAGTTGTTGT 720
 GGGCTCTTGG TGGTATCGCT CAATACGCTA TAGTGGAGAC CAGATTCTAA TACGSACAAC 780
 ACAGATTTAT ACATACTTTG TTTATAAAAC CCGAAATATG GATATGAAAC GTCTTATCAT 840
 50 GGTTTTGGST GGAGCTTCTG AATTGATCC TCAGTATAAT AAAGATGCCA CAAGCAGACC 900
 AACGGATAAT ATTCTAATAC CACAGCTAAT CAGAGAAATT GGCAGCATT AATTAAAGAA 960
 55 GAATGAGCCT CCACTTACCT GCCCATATAG CCTGAAGGCC AGAGTTCTTT TACTGTCTCA 1020
 TCTTGCTAGA ATGAAAATTC CTGAGACCCCT TGAAGAAGAT CAGCAATTCA TGCTAAAAAA 1080
 GTGTCTGCC CTACTTCAAG AAATGGTTAA TGTAACTGTC CAACTAATAG TAATGGCCCC 1140
 60

	GAACCGTGAA GAAAGGGAGT TTGGTGTTCG AACTTTGGCA TCCCTAGAAA ACTGCATGAA	1200
	GCTTTCTCAG ATGGCCGTTG AGGGACITCA GCAATTTAAG TCTCCCTTC TGCAGCTCCC	1260
5	TCATATTGAA GAGGACAATC TTAGACGGGT TTCTAATCAT AAGAAGTATA AAATTAAAC	1320
	TATCCAGGAT TTGTGAGT TAAAGAATC AGATGTCAC ACTCTACTGC ACTTCCTTGA	1380
	AGATGAAAAA TATGAAGAGG TTATGGCTGT GTTGGGAST TTCCATATG TGACCATGGA	1440
10	TATAAATCA CAGGTGTTAG ATGATGAACA TAGCAACAAC ATCACAGTAG GATCCTTAGT	1500
	TACACTGTTC GTTAAGTTC CAAGGTAAC AATGGCTGAA GTATTTGAAA AGGAGCAGTC	1560
15	CATCTGTGT GCAGAGGAAC AGCCAGCAGA AGATGGCAG GGTGAACTA ACAAGAACAG	1620
	GACAAAAGGA GGATGGCAAC AGAAGAGTAA AGGACCCAAG AAAACTGCTA AATCAAAAAA	1680
	AAAGAAACCT TTAACAAAAA AACCTACACC TGCTGTATTA CCACASTCAA AGCAACAGAA	1740
20	ACAAAAGTAG GCAAATGGAG TCGTTGGGAA TGAAGCTGCA GTAAAGGAAG ATGAAGAAGA	1800
	AGTTTCASAT AAGGGCAGTG ATTCTGAAGA AGAAGAAACC AATAGAGATT CCCAAAGTGA	1860
25	GAAAGATGAT GSTAGTGACA GAGACTCTGA TAGAGAGCAA GATGAAAAAC AAAACAAAGA	1920
	TGATGAACCA GASTGGCAAG AATTACAACA AAGCATACAG CGAAAAGAGA GAGCTCTATT	1980
	GGAAACCCAA TCAAAAATAA CACATCCTGT GTATAGCCTT TACTTTCCTG AGGAAAAACA	2040
30	AGAATGGTGG TGGCTTTACA TTGCAGATAG GAAGGAGTAG ACATTAAATAT CCATGCCATA	2100
	TCACTGTGT ACGCTGAAAG ATACAGAGGA GGTAGAGCTG AAGTTTCCTG CACCAGGCAA	2160
35	GCCTGGAAAT TATCAGTATA CTGTCTTTCT GAGATCAGAC TCCTATATGG GTTTGGATCA	2220
	GATTAAACCA TTGGAAGTTK GGAAGTTGAT GAGGCTGAAG CCGTGCCAG AAAATCACCC	2280
	ACAGTGATAT ACAGCAATAG AGGGGATGA AGAAGCAGGAG GATAGTGAGG CTTTGAAGA	2340
40	TAGCTTTGAG GGAGGAAGAG GAGCGAGGA AGAAGGTGG TGGATTAAAG GCAGTTACTC	2400
	TGGAATGGGA CCCACACTGT TTTCACCAT ATTTGGCAA TTTTTTTC CCGTTTTC	2460
45	GAGTGTCTTT CONTINUED CAGGAACCAT TACAGAACCG	2500

50 (2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 1337 base pairs

60 CTTGGCTTTC TGGGAGAGT TGGGAGAGT GTAATTTTGT TTAATTTTAA TTAATTTTAA

TCTCCCTGCG GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCGTCTTGA 120
 GCCACGGTGG CCGGCGCTGGC GGCTCCCGGG TGGTGAGAGA GCGGTCCGGG AACGATGAAG 180
 5 GCGCTGGAGT GTTGCTGCTG TCTCAGCCAC CTCTTGCTT CCGTCTCTCT CTTGCTGTTG 240
 CTGCTGAAC TAAGCGGGYC CTTGMASTC CTGCTGCAGG CAGCCGAGGC CGCGCCAGGT 300
 10 CTTGAGCTC CTGACCTAG ACCACGGACA TTACCGCCGC TGCCACCGGG CCTACCTCT 360
 CCCCAGCAGC CCGGCGCTGG TCTGCTGAA GCTGC3GGG CCGGGGGCTC CGAGGAGGC 420
 AATGACAGCA ACCCTGTGGC CCGCTTGAG ACGGACGATC ACAGAGGSA GCGCGGSA 480
 15 GCGTGGTGG GTGGGGGCTT TGCTGTGAGC CCCAACCTG CCGACAGGC CATGACCCAG 540
 CCGGCGCTGA CCGTGTGAT GTTGTGAGC GCGCGCTGC TGCTGTACTT CTTGCTCAG 600
 20 ACGGTACAGG TGAGAAGAAG AAACCGAAA ACTAGGAGAT ATGGAGTTT GGACACTAAC 660
 ATAGAAAATA TGSAATTGAC ACCTTTAGAA CAGGATGATG AGGATGATGA CAACAGTTG 720
 TTTGATGCA ATCATCTCG AAGATAAGAA TGTCCTTTT GATGAAAGAA CTTTATCTTT 780
 25 CTACAATGAA GAGTGAATT TCTATGTTA AGGAATAAGA AGCCACTATA TCAATGTTG 840
 GGGGTATTT AAGTTACATA TATTTAACA ACCTTTAAT TGCTGTGCA ATAAATACCG 900
 30 TATCTTTTA TTATATCTT ATATGTATAG AAGTACTCT TTAATGGCT CAGAGATGTT 960
 GGGGATAAAG TATACTGTAA TAATTTATCT GTTGAAAAT TACTATAAAA CGGTGTTTT 1020
 TGATCGTTT TTGTTCTCG CTACCATAT GATTGTAAAT TGTTTTATGT ATTAATCAGT 1080
 35 TAATGCTAAT TATTTTGCT GATGTCATAT GTTAAAGAGC TATAAATTC AACACCAAC 1140
 TGGTGTGTA AAATAATTA AAATTTCTT TACTGAAAG TATTTCCAT TTTGTGGG 1200
 40 AAAAGAAGC AAATTTATTA CTTGTGTTG GGGTTTTTA AATATTAAGA AATGTCTAAG 1260
 TTATGTTTG CAAAACAATA AATATGATTT TAAATTCTCT TAAAAA AAAAACC 1320
 45 CCGGGGGGG GCGCGG 1337

50 (2) INFORMATION FOR SEQ ID NO: 186:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 941 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

60

GGCAGAGCC TGGACGAGC AGCCACCGCC GCGTCCCTCT CTCCAGAGG CTGCGGCTT 60

AGGACCCCCA GCTCCGACAT GTCGCCCTCT GGTGGCTTGT GTCTTCTCAC CATCGTTGGC 120
 CTGATTCTTC CCACGAGAGG ACAGACGTTG AAGGATACCA GTTCAGTTC TTCAGCAGAC 180
 5 TCAACTATCA TGGACATTCA GGTCCCGACA CGAGCCCCAG ATGCAGTCTA CACAGAAGTC 240
 CAGCCACCT CTCCAACCCC AACCTGGCCT GCTGATJAAA CACCACAACC CCAGACCCAG 300
 ACCCAGCAAC TGAAGGAAC GGATGGGCT CTAGTGACAG ATCCAGAGAC ACACAAGAGC 360
 10 ACCAAAGCAG CTCATCCGAC TGATGACACC ACGACCTCT CTGAGAGACC ATCCCCAAGC 420
 ACAGACGTC AGACGACCC CCAGACCTC AAGCCATCTG GTTTTCATGA GGATGACCCC 480
 15 TTCTTCTATG ATGAACACAC CCTCCGAAA CGGGGCTCT TGGTCGCAGC TGTGTCTTC 540
 ATCAGAGCA TCATCATCT CACGAGTGGC AAGTCCAGGC AGCTGTCCCG GTTATGCGG 600
 AATCATGCA GTTGAGTCCA TCAGAAACAG GAGCTGACAA CCYGTGGGC ACCCGAAGAC 660
 20 CAAGCCCTCT GCGAGTCTAC CGTGCCAGC CTCTGCTATC CCTTGAAGA GCCTGGCCAG 720
 AGAGGAAGA CACAGATGAT GAAGCTGGAG CCAGGGCTGC CGGTCCGAST CTCCTACCTC 780
 25 CCCCACCTCT GCGGCCCCCT GAAGGCTACC TGGGCTCTTG GGGGTGTCC CTCAGTTAT 840
 CTCCTCTGYT AAGACAAAAA GTAAAGCACT GTGGTCTTTG CAAAAA AAAA AAAA 900
 AAAAAA AAAA AAAA AAAA AAAAACTCG A 941
 30

(2) INFORMATION FOR SEQ ID NO: 187:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 654 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

GAATTGCGCA CGAGGAGCT TGTGCTTTAA AGGAGGTGT CAAAGCATGT CTGAGCAGAG 60
 45 ACTTTGCGC TCTTTTAA TTAATACTTT AAAATAATTC ATATTAAAA TATCATATGT 120
 TTCCATAAAG AGGAGSATGT TTAAATGCCT CCAGACTACA TTCCTTTTTA TTCTTTGATT 180
 50 TTACCTGGGA GTCCAAAGTT CAATTCCTCAT AAAGCAAGCG TTTTATTTGT CACTTTCAAT 240
 ATACATCGGA TTGCTATGCT TAAGATGCAA TATGGGCTGC GGAAATAGGT TAACCCACAG 300
 TATTAATAAT AAAAAATTA TGGGCTATG GTGCAAGCA GTTCAATCT TAATATCT 360
 60 TACTAAAAAT AAAAAATTA TGGGCTATG GTGCAAGCA GTTCAATCT TAATATCT 540

GGAGGCTGAG GCAGGAGAAT CGTTTGAATC TGGGASTTGG AGGTTGTCAG TGAGGTSAGA 600
 TCGGGCCACA GCACTCCAGC CTGGGTGACA GGGTGAGACT CTGTCTCAAA NAGA 654

5

(2) INFORMATION FOR SEQ ID NO: 188:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1848 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

20 GAAACTGGAT CGAGAACCG GAGGSAAGCG AAGCGGAAGC CCGGAATGAG GCCGGACTGG 60
 AAAGCGGAG CGCGCCAGG CGGECTCCC CAAAAGCCTG CCCCTTCATC CCAGCGGAAA 120
 CCGCGGCCCC GCGGAGCGC GCGGCGCGCT CGGATTGCAG TCGCGGCGGC GGAGSAAGAG 180
 25 AGAGGCTCTC GCGAGCGGAA CGGCTGAGG CTGGAAGAGC ACAAACCGGC CCGGAGCGG 240
 TGCTTGGAGG AGCTGGTCTT CGCGAGCGTC GAGAAGGAG AGGAGCGGTT GGTGCGGCGT 300
 CTGCGAGGCC CGAGGGTTCA AGAACAAGAA GACTCGGGTG ACTCAGAAGT GGAGAATGAA 360
 30 GCAAAAGGTA ATTTTCCACC TCAAAAGAAG CCAGTTTGGG TCGATGAAGA AGATGAAGAT 420
 GAGGAAATGG TTGACATGAT GAACAATCGG TTTCGGAAGG ATATGATGAA AAATGCTACT 480
 35 GAAGTAAAC TTTCGAAAGA CAACCTTAAA AAGAGACTTA AAGAAGAATT CCAACATGCC 540
 ATGGGAGGAG TACCTCCCTG GCGAGAGACT ACTAAGCGGA AAACATCTTC AGATGATGAA 600
 AGTGAAGAGG ATGAAGATGA TTTGTTGCAA AGGACTGGGA ATTCATATC CACATCAACT 660
 40 TCTCTTCCAA GAGGCATCTT GAAGATGAAG AACTGCCAGC ATGCGAATGC TGAACGTCCT 720
 ACTGTTGCTC GGATCTCCAT CTGTGCAGTT CCATCCCGGT GCACAGATTG TGATGGTTGC 780
 45 TGGGATTAGA TAATGCTGTA TCACTATTTC AGGTTGATCG GAAAACAAAT CCTAAAATTC 840
 AGAGCATCTA TTTGGAAAGG TTTCATCTT TTAAGGCTTG TTTAGTCTT AATCGGGAAG 900
 AAGTTTTAGC CACGAGTACC CACAGCAAGG TTCTTTATGT CTATGACATG CTGGCTGGAA 960
 50 AGTTAATTCC TGTGCATCAA GTGAGAGGTT TGAAAGAGAA GATAGTGAGG AGCTTTGAAG 1020
 TCTCCCCAGA TGGTCTCTTC TTGCTCATAA ATGGCATTGC TGGATATTTC CATTTGCTAG 1080
 55 CAATGAAGAC CAAAGAACTG ATTGGAAGCA TGAAAATTAA TGAAGGGTT GCAGCATCCA 1140
 CATCTCTTTC AGATAGTAAG AAAGTATACG CCTCTTCGGG GATCGGAGAA GTTTATGTTT 1200
 60 GGGATGTGAA CTCAAGGAAG TGCCTTAACA GATTTGTGTA TGAAGGCAGT TTATATGGAT 1260

TAAGCATTBC CACATCTAGG AATGSACAGT ATGTTGCTTG TGGTCTAAT TGTGGAGTGG 1320
 TAAATATATA CAATCAAGAT TTTGTCTOC AAGAAACAAA CCCAAAGCCA ATAAAAGCTA 1380
 5 TAATGAACTT GGTACAGGT GTTACTTCTC TGACCTTCAA TCCTACTACA GAAATCTTGG 1440
 CAATTGCCTC AGAAAAAATG AAAGAAGCAG TCAGATTGCT TCATCTTCCT TCCTGTACAG 1500
 TATTTTCAAA CTTCOCAGTC ATTAAAAATA AGAATATTTT TCATGTTCAT ACCATCGATT 1560
 10 TTTCTCAGAG AAGTGGATAC TTTCCTTCG CGAATGAAAA GGGCAAGGCC CTGATGTATA 1620
 GGTTCACCA TTAATCAGAC TTCTAAAGAG ACTATTTGAA GTCCAGTTGA GTCACAAGAG 1680
 15 AAGCCTGTCT TGATATATCA TCTLAGAAGC TTCTCTGAAT ATGTGATAAT ATATGAAAA 1740
 TGATTATAG ATCCAGCTGT GCTTAAGAGC CACTAATCTC TTAATAAACA TGTGGCAGCT 1800
 TTTGTTTGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAACCTGA 1848
 20

25 (2) INFORMATION FOR SEQ ID NO: 189:
 (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 1145 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

35 AAAAAAAAAA CAGGGGAACN TTGGGGGCGG CTTTNNITTC CCCCTCCAGG CCATTGGGGA 60
 ATTCTTCAAG TTAATCTCTC TTGCTCTTG GCCAACAGGG CTTGTAGGGG GGAGAGATCC 120
 ACGATCATCA AGGGGTTGGA GTGCAAGCCT CACTCCAGC CCTGGCAGGC ACCCTCTTTC 180
 40 GAGAAGACCC GGTACTCTG TGGGGCAGC CTCATGCCC CCAGATGGCT CCTGACAGCA 240
 GGGCACTGCC TCAAGCCCCG CTACATAGTT CACCTGGGCC AGCACAACCT CCAGAAGGAG 300
 GAGGCTCTTG AGCAGACCCG SACAGTCACT GAGTCTTTC CCCACCCCGG CTTCAACAAC 360
 45 AGCCTCCCA ACAAGACCA CGGCAATGAC ATCATCTTG TGAAGATGAC ATGGCCAGTC 420
 TGCATCACT GGGCTCTGCG ACCCTCTCAG CTCTCTCAG GCTCTCTCAG TGCTGGCACC 480
 50 AGCTGYCTCA TTTCCGGCTG GGGCAGMACG TCCAGCCCC AGTTACGCT GCCTCAACCC 540
 TTGSGATGGG CCAACATCAC CATCTTTAG CACCAGAAT GTGAGAAGCC CTACCCCGGC 600
 TTTCTCTCTG CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT 660
 60 TACTGATCT AGGAGACAT GAAGAACAAT TAACTTAC CTCTCTCTCT CTCTCTCTCT 720

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

CCCTCCATTT CCACTTGGTG TTTGGTTCTT GTTCACTCTG TTAATAAGAA ACCCTAAGCC 900
 AAGACCCCTCT ACGAACATTC TTTGGGCCTC CTGGACTACA GGAGATGCTG TCACTTAATA 960
 ATCAACCTGG GGTTCGAAAT CAGTGAGACC TGGATTCAAA TTCTGCCTTG AAATATTGTG 1020
 ACTCTGGGAA TGACAACACT TGGTTTGTTC TCTGTTGTAT CCCCAGCCCC AAAGACAGCT 1080
 CCTGGCATA TATCAAGGTT TCAATAAATA TTTGCTAAAT GAAAAAATA AAAAAAAAAA 1140
 ACTCGA 1146

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

ACTCCCTCAC CCAGGTCCCA GCCCTGGGAA CCACCTACCG TGAGCCCTTT TGCAGATATA 60
 GACTCATTTT ATCCTCAGAT GGTCTTCAA GGTAGGTACT TTAGTCCCAT TTTAGAGATG 120
 AGACGATTGA GGCCAGAGGG GTGNGTAAC TTGCCTGGGG GCTCAGAGC ACAAAGGAG 180
 CCGAGGCAGG ATCTGACCTT TGTCTCTGG CCTCACTGCC CTCACTTTGG CATGACCCGA 240
 AGTTATGTCC CTACAAAGCA ATGCATGGTC CAAGGYCTT TTTATTGTAT TTTTATTTT 300
 AAGGGTCTTG TTCAAAACTG GTGTGAGCTC TGAGGAGTCC TGAACCCCTG GTGCAGCATC 360
 CTAGCATCTT GGGAGTCTTT TTCTGCCAC ACTGAGCTGG GCTCCTCGAG GGTGGGGCT 420
 GCTGTCCCTG GAAGCCTGGC AGCAGCACTG TATCGGTTG GCTGAAGCTG ACGCCGTGG 480
 GGTGCAGGGC TCCMGAATC CCGTTTGGC TGAAGGGGTT CCTGTAGCC MGGATGTTT 540
 ATGAGGTCTC TCTGATGCC CAGGCGCAGG ACATGTGTG GGTGGAGAA AAGCAGCCCC 600
 TTTCACTGCC AGCTCCACTC AATTCTATG TGGACCAAGA ACGATAAACT TAAAAATTT 660
 TTTTCTCTAA GGTATCTTCA GAATATGGTG TATTTTATG TGGAAAAGAA AAGTTATGAA 720
 GGCAGCTGTT ACTTTAAGAG AAAATTCATT AAAACTCCTC GAGGTATGAA GATGACGGCG 780
 TGCTTCTCAA TCATTTTGGC ATAATTGAT TGTGGCTGTA ATTTTTTTTT TTTTTTTTGT 840
 CAAGCATGTC AGACAATAAA GTCTTTGTAA AAAGRGAAAA AAAAAAAAAA AAAAAAAAAA 900
 ACTCGA 906

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

CTTCAGCTGA AGCCAGGGA CCCCTTTTCC ACCCTGGGCC CCAATGCCGT CCTTTCCCGG 60
 CAGAGACTGG TCCTGAAAC CTCAGCAAA CTCAGCATCC AGGACAACAA TGTGGACTTG 120
 15 ATTCTGSCCA CACCCCCCTT CAGCCGCTG GAGAAGTTGT ATAGTACTAT GGTGGCTTC 180
 CTAGTGACC GAAAGAACCC GGTGTGCCG AGATGGCTGT GGTACTGCTG GCCAACCTGG 240
 20 CTCACAGGA CAGCTGGCA GTCTGTGCCA TTCTACTGCA GAAGGGCAGT ATCGGCAACC 300
 TCTTGGCTT CCTAGAGGAC AGCTTGGCG CCACACAGTT CCAGCAGAGC CAAGCCAGCC 360
 TCTCCACAT GCAGAACCCA CCTTTGAGC CAATAGTGT GGACATGATG CGCGGGCTG 420
 25 CCGCGCGCT GTTGCCTTG GCCAAGGTG AGGAGAACCA CTCAGAGTTT ACTCTGTAG 480
 AATCAGGCT GTTGACATC TGGTATCAC GTTGATGAA CTCAKTGGT TCACAAGTCA 540
 30 TTTGTATGT ACTTTTTTG NATGCGCAG TCATGACAGC CGTGGGACAC CTCGCCCCC 600
 CGTGTGTGTG TGGTGTGTG GAGAACTTAG AAAGTACTG TTCCCTTTA TTATGCAAA 660
 ACCACCTTAG AATTCAGTTT ACCCTGTGCT GTCCAGCTTC TCCCTTGGGA AAAAGTCTCT 720
 35 CTTSTTTCTC TCTCTCTCTT CCACCTCCC TCCCTCCATC AACTCAGCC TTTCTGTTCC 780
 TTCTCTCAC CTACTCCCC TCAGGACCTT ACCCCACCTT CTITGAAAAG ACAAGCTCT 840
 40 GCTACATAG AAGACTTTTT TTATTTAAC CAAAGTACT GTTGTTTACA GTGAGTTTG 900
 GGAAAAAATA TAAATAAAA ATGGCTTTCC CAGTCTTCC ATCAACGGGA TGCCACATTT 960
 CATAACTGTT TTAATGTA AAAAAAAAAA AAAAAATAC AAAAAAAT TCTGAAGGAC 1020
 45 AAAAAAGTG ACTGCTGAC TGTGTGTGT TTATGTTGT ACATTCACAA TCTTGCAAGA 1080
 GCCAAGAAGT TGGCAGTTGT GAACAGATCC TGTTCAGTGG AGAGGCTGT GCAGTAGAGT 1140
 50 GTAGACCTTT TCATGTACTG TACTGTACAC CTGATACTGT AAACATACTG TAATAATAAT 1200
 GTCTCACATG GAAACAGAAA AGCTGGGTG AGCAGCAGGC TGTAGTTTTT AAAAAATGTT 1260
 GAGTGTGAC ATTCAACATA GCTTGGATC GCGAATTTT GTCTGCTT CCACTACTG 1320
 GAGTGTGAC ATTCAACATA GCTTGGATC GCGAATTTT GTCTGCTT CCACTACTG 1380
 60 GAGTGTGAC ATTCAACATA GCTTGGATC GCGAATTTT GTCTGCTT CCACTACTG 1440

5 CTTTATAGTA TGACGAGTTA ACAAGTTGGT GACCTGCACA AAGCGAGACA CAGCTATTTA 1560
 ATCTCTTGCC CAGATATCGG CCTCTTGGT GCGATGCTGT ACAGCTCTCT GTAAAAAGTC 1620
 CTGCTGTCT CAGCAGCCAA TCAACTTATA GTTTATTTT TTCTGGGTTT TTGTTTTGTT 1680
 TTGTTTTCTT TCTAATCGAG GTGTGAAAA GTTCTAGGTT CAGTTGAAGT TCTGATGAAG 1740
 10 AAACACAATT GAGATTTTTT CAGTGATAAA ATCTGCATAT TTGATTTTCA ACAATGTAGC 1800
 TAAAACTTGA TGTAATTCG TCCTTTTTTT CCTTTTTTGG CTTAATGANT ATCATTTATT 1860
 CAGTATGAAA TCTTTATACT ATATGTTCCA CGTGTGAAGA ATAAATGTAC ATTAAATCTT 1920
 15 GGTAAAGACTT TAAAAAATAA A 1941

20

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 2118 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192

30 AAATAATAAT AANAATAAAT AAAAATWAAG TGCTTAKTGT AACTCAGCGG ACAGGGCTCC 60
 CAGCTGCTCT GGCACGTGGG ACACCTTCCA CCTGCACAC AACAGGCATG CAAAGAGGAC 120
 35 TGGATATGGT GGGGTAGAGT GCTTCTGGTG TGTTCACTTT AAGAAAACAT CTGCCAAGAG 180
 AGAAGAGTGC CCAGGAAAGA CCAGGAAAT ACAAGTACAT GCTGCTTCA TACCATATAC 240
 40 CCCAATTCTT TAAAGCAGCA AAAGGCACTT TTTTTCAG GGCAGAGTGA ATCTAAAACA 300
 AACCTGGCTT TGCTTACAGG GAAGCTGTCC CAGAAGGACT GAGTGATGCC TCTTGTTCCC 360
 TAAGGTCTGG AGAGTCTTTG CAAGTTTCCA ACGACATTTT CAACCAGGTG GSAGAGACCA 420
 45 GCAGTTGACG AGACAAGTCA GACCCAAAAA ACGACGCCAA GGTAGTGACT GGGTGCCTAT 480
 TTGGGAGTAG GATGATTTGA GGAAAACAGG AAGAAAACC GGTGAGAAAG TGGCACTTTG 540
 GAAGTGAAA GCTGTTTGCA AATAGCAACT CTGGCTAAAG CGAAAATGTT AATCAAGTAG 600
 50 AAAGTAAAT TCAGGATCTT AGAAGCTCAT CCTTCTGATG AGAACTATTT TTTTTCCTG 660
 GAAGGAACCT TTATTACTTT AAAAGTGAGG GTAATTTACA TATGGGGTGT ATATATTCTA 720
 55 AAAATAGTAA TAAAGTACC TTTTATAAGC AATGTTGTGT GGCTGTAGA AGAAAGCAGG 780
 GAGGAAAAAA AGGCAGGCAA AACTAGTCTA GGTCTAGGCC CTAAAAATGA GCTTCCTTCC 840
 CACTTGACTG GAAACGCCCA TGTGATTTCT AGGCTGAAAA TAGGTAGGAT TTAACGAGTA 900
 60

ACCTAGTTCC CTCTGTCTC TGATTTCTGA TCAGCTGATG GAGCTGCTAG TAAGAGGGG 960
 CGATCATGCT CCCAGACGAG TCTTTTGGCC TCTTGCTCTC CATCCCAAGC CTGACTCCTT 1020
 5 CAGCAGCAGC CCCTCCTTC TGTGTCCATC TGATGCAGGC AAGCAGGAGC AGTAAGAGGG 1080
 CATCCCATGT TCCAGTTTAC CTCTATGGG GTGACTARGA GGTTCCTGGT AACTAGGGCA 1140
 GCCCARGGCC AGCAGGTTGC AFAAGCAGCT GCAAGCTTCA GAAACCCACT TCTTCCAACA 1200
 10 CCACGGAGST GCCAGAGAGC CCATCCAAAA GCCCACTGGG AGAGCCATAA GATTCTGTGC 1260
 CAGGCCCCCA GGTCCCTCTT GTGTCAAGTA GGCTCTGCTA CTGGCTCTG AAGTAAAGGC 1320
 15 AAANACAAAC GGCAGGGCA GGGTGGCAGG AATAAAAAAC TCTGACAGA AACCTTTTA 1380
 ATAAAGGAAA TTCCACCCCT CCAATCTTT CCATGGAAGG GTGAGACCTT AATGTGATGT 1440
 AAGAGGAAGG TTTTCTCTGG CTTTCAGGGA AACAGCTGCA GCTGAAACTT AGGCGGCCAT 1500
 20 TCCAGGGTAC TTTTCAACCAC AGCCAGTGCA GCGCTCCAA GTGCTACTGT CAGCCCCATC 1560
 ACTGCCAATT TCAAAAGGG GTTGGTCTTT GGCTTGSTCA GACATCTTT TGTGATCT 1620
 25 TCAGGCCCA GAAGTCCCG AANATCGCTG CCGCAGTACC ATATCAGGCC TGTGCTGGGC 1680
 TGATGCCAGC TCAAAGTUTT TGAAGTAGA GGCTGCGTTC CTCTAGCTT GCTGTTGGGC 1740
 AGGGGCTCTC CGAGCAAGTT CGGATGGGGG AACTGAACA AAAAGGTCTC CTSTCTGCTG 1800
 30 ATTAGTCTCT CATACGGTAA STCCTGAGGG ATCTGGACA ADAGTGGTG GATGAGGCC 1860
 ATGTCAAGT CACAGTCCAG GACTTCTGTC TCGGATACA ACACAATCAC GGCTGCAAAG 1920
 35 TAAATCGCA TCASTGGTG GCAGGCCAGG AAGAAGTCAT ATAACGCAC GACGTGCTG 1980
 AAGTCAGACA GGACATGCC AAACCAGTG ATGAGTCAGC TGAGGCAAA GATGCTCCT 2040
 40 AACTCAGCAC TCTGATGAA GTGATGGAGC TCTGATTCA CCTGTCAT GATGGGCATC 2100
 AGATAGTTTA ATATATGC 2118

45

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 1538 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

ATCAAGGCT GCTGATGGG GCGCGGAGC TGAAGCTTC TGTAGTCTT TAAAGAGA 180

GACGGGAGG CACCTCTCGT CCCGAAACCG ACCAGAGGGG AAAGTGITGG AGACAGTTGG 240
 TGTGTTTGAG GTGCCAAAAT AGAATGGAAA ATATGAGACC GGGCAGTTTT TCCTTCATAG 300
 5 CATTTTTGGG TACCGAGGTG TGGTCTGT TTCCCTGGCAG GGCAGACTGT RTGACCGGGA 360
 TGTGGCTTCT GAGCTCCAG AAAAACAAGA GAACCTGCT GGCATGGCT CCAAGGAGGT 420
 10 GAAAGGCAAA ACTCACAATT ACTATCAGGT GCTGATTGAT GCTCGTGAAT GCCACATAT 480
 ATCTCAGAGA TCTCAGACAG AAGCTGTGAC CTTTGTGGCT AACCATGATG ACAGTCGGGC 540
 CCTCTATGCC ATCCCAGGCT TGGACTATGT CAGCATGAA GACATCCTCC CCTACACCTC 600
 15 CACTGATCAG GTTCCCATCC AACATGAACT CTTTGAAGA TTTCTTCTGT ATGACCAGAC 660
 AAAAGCACTT CTTTGTGTGG CTCGGGAGAC GCTAAGGGGC TGGCAAGAGA AGAATCACCC 720
 20 CTGGCTGGAG CTCTCCGATG TTCATCGGGA AACAAGTGA AACATACGTG TCACTGTGAT 780
 CCCCTTCTAC ATGGGCATGA GGAAGGCCA GAATTCCTAC GTGTACTGCT GCGCTACTG 840
 TATCCGTTTG GAGAACCTTG ACAGTGATGT GGTACAGCTC CGGGAGCGGC ACTGAGGAT 900
 25 ATTCACTGTC TCTGGCACCT TGGAGACAT GCGAGGCCGA GGGGTAGTGG GCAGGGAACC 960
 AGTGTTATTC AAGGAGCAGC CTGCGTTCCA GTATAGCAGC CACGTCTGCG TGCAGGCTTC 1020
 30 CAGTGGGTAC ATGTGGGGCA CGTTCGGTT TBAAGACCT GATGCTCTCC ACTTTGATGT 1080
 TGGATTCTT CCGTTCTCC TGGAAAGCAA TAAAGATGAG AAGACACCAC CCTCAGGCT 1140
 TCACTGTAG GCCAGCTGAG GCCCAACTG CCCAGGTTG GTCACCGGGA AGAACAATC 1200
 35 TCATCCACA ATTGCTGCAG AACTCTTCTC TCCCATCAT GGGCCACAGT GGGTCTCTTA 1260
 ATTTGATTGT GGGGTTCTTT TTGTTGGGAG GGGTGGTATA ACTTTTCTTC AGAAGACCCA 1320
 40 TGTGGGACAC CTCCAAGGCT GGCTCCTCA TAAGCCCTGC CTACACCATG TTCCAGTAAA 1380
 CCTCTCCACC AAGGAAGTGT GTTCACTGC CACAGGCTG GAGGAGTTTC CTGGCCTGTC 1440
 45 ACGTGAGGTT TGATCAGTAA ACCAGTGCAS GYTTGGCCAA AAAAAAAAAA AAAAAAAAAA 1500
 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCGA 1538

50

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS.

55

- (A) LENGTH: 1098 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

60

AGACCCCTGTC TCAAATAMTA ATAATAATAA TAATCTTATT TTGGAGAATA AAGAGACCTS 60
 TGGATTTCAG GTGCCATTTC GGTAGAAAGA AAAGACGTTT ACACCGAGAA ATAGTCTGTG 120
 5 TTGCCCTGAA GGAGCAGAGG GATGCATGCC TGGAGGTGAC CTACAGTTGA AGAAGACTCA 180
 TTATGACAGA CTTGTCTCTT CTTCCTTSTG GAAASTGTTT CCTCTGCTCC TACTGCTCAT 240
 GAGACTCTTC CCCCTCCTTG TCCAGGGGAA CCAAGGGGCT TTNCTACCAC ACCCTTTCTT 300
 10 NGCCCCCGGC CTCCCATGTC TGCTGTGCTT TTGTACTCAG CAATTCTTNG TTGCTCCCA 360
 TTATCTTCCA SCCGGATACA GASTGAATAG TTAACCACAC TTAGGTCAAA TAGGATCTAA 420
 15 ATTTTGTTC CTGCTCCNGT GTAAAGAGGC CAGTGTTTGT GTGTTGGAAG CAGCCTTCCA 480
 ATAGTAACTC TTCTCATTTG TTTGGGATCT CGGCAMCAAG TTCCAGAATG ATACACGGAT 540
 CAGTGCAGAA GTTCATCAGG CTCTGGGACC TTAGGGCTGT TGGAGAAGGC TTACGACCA 600
 20 GAAATGATGG TKAWKGYTCG TGTTCTCCAT CCTCAACTTT CTTTGTCTCG ATCATAACA 660
 AGAATACATT TGAAGGGCA AAAAATGAAC ACTGTTGTTT ATTGAGCCG TGTTTTGTGA 720
 25 CACAGATGCA CAGTCTGCTG TGAAGAGTTT CTCTCAAGTG GSATYTGGGA GTCCATGCCA 780
 GATCATGGTG CTTTCATGAG GACTGACAGC TATCAGGGGT TGTGGCACTT AGTGAGGACT 840
 CTCTCCCCC AGTGTGTGCT GATGACACAT ACACACCTGA CAATAGCTTG ACTCTTCTCT 900
 30 GTTCTTTTTA CTCTGTAGCC AACATACACA TGAATTAATA CCCTTTCTAA ATATCTATCA 960
 TGGTTCATCT TTGTCCAAAT GCAGAGTCAG AGTATTTTGT ACTTCATTAT TATTTCGAAG 1020
 35 GCGAATAGTT GGCCTTCTTT TTGCAAAAAT AATTAAAGTT TTTGTATGTT GCAAAAAAAA 1080
 AAAAAAAA CTACGTAG 1098

40

(2) INFORMATION FOR SEQ ID NO: 195

45

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1001 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

GAATTCGGCA GGAGATACCT TGCATCTCAT CCCAGTAAAA CCACTTATTT ATACATATC 60
 GAGTTAAAT ATCTCTCTT AATCTTTT TTTTCTCTT TTTTCTCTT TTTTCTCTT 120
 GAGTTAAAT ATCTCTCTT AATCTTTT TTTTCTCTT TTTTCTCTT TTTTCTCTT 180
 60 ATACTCTCAT CACTGAGAT CAACTCTCA GATTAATCA TAAAGATTT CTAGAGACA 240

AGGCCAGGCC TSATCCCTGA GGGATGCATG AGAAGGCTTG GAATTCATTT CTGCTATGCT 360
 GGCTCTCTCT TSATCTTTT GAGTAGCAAA AAACAGCAAT GTGGGCCCAA TGGTGTGCCC 420
 5 TAAATCATCA CAAAGGTAAA TSAGTAAAGG GCTCAGCAGA TSAGTAAAGG GGCCTGTCTT 480
 GAGAAATTAG CACTGGGCTC TGCATTCAGA AACATGTGAT AAGCATTGCC CATTGCACAT 540
 10 TGGTTTTATT GTTAAGGAC ATGAAATTC AGTTTTGCAT AGCTAGTGAT GAATACTGA 600
 AGGSAATTGC AGACATATTT TATTTTATTT TTAATTGACA GATGGAATTG TATATATTTA 660
 TCAATACAT AATCATGCTT TAAAAATGT ACATTTATGA APEGCTAAAT CAAACTAACC 720
 15 TAGGCATTAT CTCATATAAT TGTCATTTTT GTGCGAGAA GACTAAAAAT CTACCTTTTC 780
 AGCATTTTTA AAGAATACAA TGTGTTTTAT TAACAACAGT CACCATTTGG TACACTAGAT 840
 20 CTCTTGAAGT TCTTCTCTTT ATCTAACTGA GATCTTGTAA CTTTGTATAA CAGCTCCCAA 900
 GCCCTTCCCG AAGCACTGCT CCACCGCTGG TAAACACCAT TCTATTCTCA ACTTCTCTGT 960
 25 AATCACCATT CTAGACACAG GGAAGACTCT CTACCTCTG A 1001

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1443 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

ATAAACTGAA ATAGGTCATG CAAATATAAA ATATTATTTT TAAATTATTT GTCATAAGAA 60
 40 ACGATGGTGG CCATATTTTG CTTTAATAAT GGAAAAATG TGGTTAGCAT TCTKTGGAAG 120
 GTGGTCATCA GATAGTAGAC ATTTTCTAGG ATTTATTTCT ACCTGCATAT GTGGAAATGT 180
 45 GTACTACTTT AGATTTATWT AATGGCAGCT AACTCAGAGG CATCAAAATG TGCTAATGGT 240
 GTAATATGSC CTTTGTCTTG CTGTCTGTT TTGTARGCT TCAATCAAGC ARGGGCAGGG 300
 CCCTACAGTG AACTTGTCTT TTGSCAGAG CCAGCGTCTG CCCCTGACCC CGTCTCCACT 360
 50 CTCTGTGTCC TGGAGGAGGA GCCCCCTGAT GCTTACCTG ATTCACCTTC TGGGTGCTTT 420
 GTACTGAACT GGAAGAGCC GTGCAATAAC GGATCTGAAA TCCTTGCTTA CACCATTGAT 480
 55 CTAGGAGACA CTAGCATTAC CGTGGGCAAC ACCACCATGC ATGTTATGAA AGATCTCTTT 540
 CCAGAAACCA CCTACCGGTG AGTSCAAGG AGTAGAAATC TGCATCAGCA CATCAGCACT 600
 TGGGGATCTA AGTAAACCTC TCGGGGAAAA TGAACAAGTG GATGTCATCT CCCAGCTGTT 660
 60

TCTAAGAGGC CAGATGTCCA GASTATTSTC TCACCTTCAT CCGTCAGGTC ABAATACCTG 720
 TGAAAAAGCC ACACTGGTTC AGGGAAGTAC TGGAGGCTTT TGTGTCCAT ATACCTTCCA 780
 5 CCGTCTCTAC CCCAGAGTGG ACTCARATCC TCAAGTATG CTCTGAGCAT TCTGCTGAGA 840
 AATTATAAAA GGGCTTTGGC AATATGTTAG CCGAAGATT TGGCTTTTC CAAAAATTGT 900
 GCGGACNCTA ACAGTGGCTT AAATGATGGT AAAACTTTTA AGATTCTTAA AAGTCTGGCA 960
 10 TTGGAGATAC GTTGACTTTT ATTAAGACMAC CTATATTTGT TTAATCATTT CTAAGAAAAA 1020
 ATCTGGAGCT CAGGGGTTC ACTGACGGAA GACATTTTGA CATCATTTCT TAACTAATTA 1080
 15 AATGCCAGXT AACCCGTGGA AATTATCAAA AACATCTTTC AGTACCTAGA AAGGACTTGA 1140
 GAGGATAGTT CTGTTATGGA GAAGATGAAA TGGTTTACTA GTGTAGGAC TATGAAAGG 1200
 TGAGCTTAGA TTGGATAGT AAAACCTCAA GACCTTATTT AAAAGTATT TTAAGATGC 1260
 20 AGCATAAATA ATTTAATTC GTGTTAANAT GCGAGCTTA STATATTTAG CTGATTTGA 1320
 AAAGAAATC ACATTGGGAG AATGCCACCT TTTCTTATA AGATAGCTTT GAGATACCA 1380
 25 TTTTAGACAG ATGGAAATG AATAGCTTTA GAAAGGCA ATGTTTGA TCCTGGGAAA 1440
 AAA 1443

30

(2) INFORMATION FOR SEQ ID NO: 197:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

GAAAAAAGAA AGTATGAGCC AGTAGCTAGG CAGCTGTGGC CCGGCGACT TCAACATTA 60
 AATTAAGTGT CACAGTATCA TCTTAGAAGT GAAGAAAGC CTTTATCTT GAATGGGCG 120
 45 TCTACACCA CTTACTGACA AAGAACATG TCTATCTTG CATGGGAAA ATCTTCAAT 180
 TCTATGGCT TGTATGTGTC CCGTCAAAAT CAGTCTTTC CAATGTGACA CCAATGAGG 240
 50 GTGGGCTCTT TAAGAGATCA CTAGGCCATG AGGGAATCTC TTACGACTGG CATGAAGGCC 300
 CATAATAAAA GAGCTTTTCA GAGGATCTCT CCTATTTTC CTCTGTATG TGAACACAA 360
 GAGGATCTCT CCTATTTTC CTCTGTATG TGAACACAA 420

60

AATCTTTA TTGATCTC AGTATTCTT ACTTATTAAG GAATGAGAA AAGTATCTT

AGGGCATAGG ATGAACAAGT TACTGCTAGA COTCTCAGAA TGCCACTAAT GGTATAGATT 660
 GTATTTTCAT CATTTTGTGT CTCTTCGGAA GGTAAACAGG TCTATATAA GGTACTAAT 720
 5 AGATGTCTAA AAACACCTTA AGTATTTGTC TAAAAATCTG CTGCATTCTG GAGAAAGAG 780
 CAAAATTCMA AATAATTTCA AAGGGCCTAA AGGACTATTT AATGAAATG GATTAATTTT 840
 10 TAATGGTACT ACCACTCTCA AATTTAAAT GTTATCTTAC GTTCTCTCTG CTGGATTTGG 900
 ATTTATTGCT AAAACCTGCT AAACACTTTA ATCTTTTCA ATCCATTAC GATCTCTCTT 960
 GTCCAGAATT ACTCCAGAC TAATAGTCAC CTACTTCTG CTTCTGATC GCGATTTGCT 1020
 15 GTTAATTTCT GGTACAAAT AACTAACTGC CAACTAATC TTCTAATAA GGTAACTGA 1080
 TCTGCTCACT CTTTGTCTCA ACAATGTAAA ACCTCCGATT GTCTCCGAA TAAACAGAG 1140
 20 TTCCACTCT GTATACAATA CATCCATGAT CTGTATCCAG CATCATTTTG TATTGCTCA 1200
 CTTTATACAG CACCCCCCAT GGCACATCAA ATTAATTAAT CTTGATTAAT GGTACTGCA 1260
 25 AAAAAAAAAA AAAAAAATC GA 1282

(2) INFORMATION FOR SEQ ID NO: 198
 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 951 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198.

ATTTGGGAAC GAGGACTGAA GTGGGAGCGG CCGCAGGTA GAGACAGAA GGGGATCTA 60
 40 TGTGGTAACT AAAGAATGTT TCTGTTTTGT TAATTATTGT GTGTCTGTG TTTTATGTT 120
 TGCTTAAGAG AATCAAAAAC TGAAAAAAT GAGAATACAG GAAATGGCTC TTGTTTATTT 180
 45 TTTTGTCTGT TTTACAGCTT GTTAATGCTC TACTGTCTTT GTTCAGAGAG AGATTGTTC 240
 ACTGCCCCAGC TCGTTTTGTG TCTGAGCCC TATGCCAGC CCACTTATA AATCATGCTT 300
 GTTAGATGT TTGATTTGT TCTGTTTGT ATTGTTATCT TAAAGGTGTA TAACTCTGAC 360
 50 ATGCCAGACA TCAAATTAAG CTCAAATTAA GCTCTGTTT AATGTTTTAA ACACCTAAT 420
 TATATTCTAA TTGATCCAG CCACTGATGC ATGTACTTTA GCTACTTCTG CTAAATAAGC 480
 55 ATATTAATTT TCCACATCAG GCCATCAGAT CTTGAGAACC AACAGTTATC TAGAATTCG 540
 TGTCTACTAA TGTTCACCT GCATGCAGCC TTCAATTAAT TTGTACAAA ATATAAGTG 600
 60 ATCATTATGT AGTTTCTGGA TTAATAAAT TTGTGTGTGA AGTTGCTTTG TAAAGTCAT 660

GTGGAATTAA TGGGACAGTG TGCCCTTTGT GTTAGATGTT AGAGCAAAAG AAAGGGCTTA 720
 TAGTGTAGT ATTGGAGCAC TTTGAAGATA GATATTTTCA GAAAAGATGT AGGATTTAAA 780
 5 AGTTAAATTT TAAATTTTAG AAAAAGATAT GATGGCAAT CGAAATAGTC ACAATGAAGT 840
 TCTTCATCCA GTAGGTGTTT AACAGTGTTA TTTTGCCACT GGTAAATGTCT AAAGTGTGAG 900
 TGATTTACAA TAAATGATTA TGAATTCAAA AAAAAAAAAA AAAAACTG A 951
 10

(2) INFORMATION FOR SEQ ID NO: 199:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1740 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

TTATTATAAT AATGATGATG ATTCCAAGGA AAAAACCTAC AGCGAATGTT CCATTTCTAC 60
 25 CCCGCACGCA GACACTCTCT CTAACACTGA TAACCTGAGC CCCAGCACT GGACGGAAGA 120
 ATGCTGCGCT CTCCTGTGT ACTGGTTCAG GGTTCGGGC CCAGCCTTGT CAGGACCCCC 180
 30 TGGTGTCCAG AGCCCCCACC CCTCCCGCAA CAAGCAGCTG ATGCCCCAGT GATTCTCTAT 240
 ACATTTTTC CTTGGGCCAA TATGTCCAGG AAAACTGCTT ACTTCTCTTT TCTTGCTCG 300
 AGCCTTCATT GTTCACCCCT ACGTTGCAAT ATAGGAATTA ATGCTACAAA ATAAAAAGTAA 360
 35 AGCTTACCTG AAAAGTGCAT AGTTTGGGGC AATGGTATCT ACATCTCCCA CTGTGGGAAA 420
 ACCAGCAAAG CATCAAAACT CTCAATTCTC CTGTTACCFA ATGCAGATCT GAATTATAAG 480
 40 ATGTTTATGT TTGACCATG TTCAACAAT GGGATTTTGT TACGAATTAT CCTTTAACT 540
 GAAACCTCA GTTTTACTGT TTACATTAT AGGAAAACAG GATATTTT TGAATCTAAA 600
 AATTTGATGT ACAGCATGTG ATTTTGAAG TTTACATGTA AAGTCAAGT ATAGGTGAAA 660
 45 TAAGCTTGT CATATTTTGA GAGGTATCCT GCAGCATGT TTTTACGTGA GTGTTTGTAGT 720
 CAAATACAT GGTAGACAGT CTTTCACAAT AAAAGGAAAA GGATTTTMTT TCCTCCAAAT 780
 50 GTACATTTAT CAAGCTAATG ATTGATTTT TTAAAAAGAG ATTTGCCCC AGTCTGTTT 840
 ATGAAAGTTC ATTGCCCTAA ACTGTGCTGA TTGTTTTTAA TCAAGTTATA AATTTCCAAC 900

60 AATGCAAGCT TTAAATTAT ACTTCAACT GAAAGATGT CTTTCTTA CTGTTTCTT 114

CAGGAACTTT ACTTCAGAGC TGTCCAGATT GCAGTTGTGC CCGGTGTATG TGGATCTAGT 1200
 TCACAGAGTC TTTCGAAAGCC AAGCACTCGTG CCGTCCGTAT ACTGTCCACT CATTTTATGT 1260
 5 AGATTTCGTA TCCTCAGCAG CCACTGTTAA CAGCACTGTC AGTACTTAN CAGATTCATC 1320
 TTTTATGTAT TTAAAGTAAT CCATACTATG ATTTCGTTTT TCCTGCACC ATTAATTCTG 1380
 10 GCATCAGATC AGTTTTTCTG TTGTGAAGTT CTACTGTGGT TTGACCCAAG ACCACAACCA 1440
 TGAGACCCCTG AAGTAAAGAT AAGGTACACA TATATTATTT GAGTAACTGT TTCTTGGGG 1500
 GCCAATCTGT GTATGCTTTT ABAAGTTTAC AGAATGCTTT TATTTTGTG TATAACAAAC 1560
 15 AGTCTGTGAT TTATTCTGT TCATAAACCA TTTCGACAGA GTGAGGACGT TTGCGCTGTT 1620
 ATCTCCTAGT GCTAACAATA CACTCCAGTC ATGAGCGGCG CTTCACAAAT AAAGCAUTTT 1680
 20 TGATGACTCA MAAAAAAAAA AAAAAAAAAA YCGCGGGGGG CCGGCTAACT CATTTCNCCC 1740

25 (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1707 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

35 GCTTATAGAA GGGAGAGGAG CGAACATGGC AGCGCGTTGG CGGTMTTNGT GTGTCTNGT 60
 GACCATGGTG GTGGCGCTGC TCATCGTTTG CAGCGTTCCG TCAGCCTCTG CCGAAGAAA 120
 GAAGGAGATG GTGTTATCTG AAAAGCTTAG TCAGCTGATG GAATGGACTA ACAAAGACC 180
 40 TGTAATAAGA ATGAATGGAG ACAAGTTCG TCGCTTGTG AAAGCCCGAC CGAGAAATTA 240
 CTCGTTATC GTCATGTTCA CTGCTCTCCA ATGTCATAGA CAGTGTGTG TTTGCAAGCA 300
 45 AGCTGATGAA GAATCCAGA TCCTGGCAAA CTCTGGCGA TACTCCAGTG CATTACCAA 360
 CAGGATATTT TTGCGCATGG TGGATTTTGA TGAAGGCTCT GATGTATTT AGATGCTAAA 420
 CATGAATCA GCTCCAACTT TCATCAACTT TCCTGCAAAA GGGAAACCCA AACGGGTTGA 480
 50 TACATATGAG TTACAGGTGC GGGGTTTTTC AGCTGAGCAG ATTGCGCGGT GGATCGCCGA 540
 CAGAAGTAT GTCAATATTA GAGTGATTAG ACCCCCAAAT TATGCTGGTC CCCTTATGTT 600
 55 GGGATTGCTT TTGGCTGTTA TTGGTGGACT TGTGTATCTT CGAAGAGTAA TATGGAATTT 660
 CTCMTTAATA AAAGTGGATG GGCTTTTGCA GCTTTGTGTT TTGTGCTTGC TATGACATCT 720
 60 GGTCAAATGT GGAACCATAT AAGAGGACCA CCATATGCCC ATAAGAATCC CCACACGGGA 780

	CATGTGAATT ATATCCATGG AAGCACTCAA GCCCAGTTTS TAGCTGAAGC ACACATTGTT	840
	CTTCTGTTTA ATGGTGGAGT TACCTTAGGA ATGCTGCTTT TATGTGAAGC TGCTACCTCT	900
5	GACATGGATA TTGSAAGCGG AAGATAATG TGTGTGCTG STATTGGACT TGTGTATTA	960
	TTCTTCAGTT GGATGCTCTG TATTTTGA TCTAAATATC ATGGCTACCC ATACAGCTTT	1020
	CTGATGAGTT AAAAAGGTCC CAGAGATATA TAGACACTGG AGTACTGGAA ATTGAAAAC	1080
10	GAAAATCGTG TGTGTTTGAA AAGAAGAATG CAACCTGTAT ATTTGTATT ACCTCTTTT	1140
	TTCAAGTGAT TTAATAGTT AATCATTTAA CCAAGAAGA TGTGTAGTCC CTTAACAAGC	1200
15	AATGCTCTGT CAAAATCTGA GGTATTTGAA AATAATTATC CTCTTAACCT TCTCTTCCA	1260
	GTGAACTTTA TCGAACATTT AATTTAGTAC AATTAAGTAT ATTATAAAAA TTGTAAACT	1320
	ACTACTTTGT TTATGTTAGA ACAAACTCA AAACTACTTT AGTTAACTTG GTCATCTGAT	1380
20	TTTATATTTC GTATCCAAA GATGGGAAA GTAACCTCG ACCAGGTGTT CCCACATATG	1440
	CCTGTTACAG ATAACATACAT TAGGAATCA TTCTTAGCTT CTGATCTTT GTGTGGATGT	1500
25	GTATACTTTA CGCATCTTTC CTTTGTAGTA GAGAAATTAT GTGTGTCATG TGGTCTTCTG	1560
	AAAATGGAAC ACCATTCTTC AGAGACACG TCTAGCCCTC AGCAAGACAG TTGTTTCTCC	1620
	TCTCTCTTCC ATATTCTCTA CTGAAATACA GTGCTCTCTA TGATTGTTTT TGTPTTGTG	1680
30	TTTTTTCAG ATCAGGYTAC TGGGCTC	1707

35

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 779 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

45	CTGTCCCGAG TGTTCCAGG TAATGACTTG GCACTCCAGA GAAAGTTTCA TGTGTTGCG	60
	TGTGTTGGCT CCAAGCCAAG CACCTGGCAT GCAGGTGAGC CCTTCCAGC GGGCGTGCG	120
50	TGCTCTCTTT CACAGATGCC ACSTTGCAGC CCAAGGCTT CACCATTTTG CGTTTTTTAG	180
	AAAGCCATTT TTTTGGTCTT TTATAAAGCT GCTTATAGA TATCTTTGAT CTGTCATGC	240

60

TACTTAAAT ATTAATAT ATTTCTTTT TATCTAT AG TATAAAT TTTTATTA

TCTAATAT GTTTTATTT CACAGAGAGC TCGGAATAT CTCTTATAG TATACACTG

480

454

TAGGTTTCAT AAATTTTAAG AAAGTTGCTT TTAAGTACTA TTTATAGGTT TTTCTGTTAT 540
 ACTTGCAACC TAGTTTTAAA ATACATGAGG ATTTTATGAA AGCTTTATAC AGACATTTAT 600
 5 AGGAAACTCA TTTTGTGATT TTAGTGTCCA TTAAATTGA TAACACTTAC TTTATAAAAA 660
 GATGCTTTTT GTCTGGATAG AGCTTATAG TTTAAATAT CTTCATATAT TGCCATTTGA 720
 10 TCAAATAAAT TTCTTACTTA GAAAAAAAAA AAAAAAATAA AAAAAAAAAA AAAACTCGA 779

15 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1617 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

25 GGCACAGCTT TCTGTCTTCT CCTCGCTCCC TCTCTTCTC TCCTCCCTCT GCGTCCCAG 60
 TGCATAAAGT CTCTGTGCTT CCGGAAGCTT GTTGGCAATG CCTATTTTTT GCGTTTCCCC 120
 CGCGTTCTCT AAACATACTA TTTAAAGGTC TGGGTGTGCA AATGGTTTGA CTAAACGTAG 180
 30 GATGGGACTT AAGTTGAACG GCAGATATAT TTCACTGATC CTGCGGTGTC AAATAGCGTA 240
 TCTGTGTGAG GCCGTGAGAG CAGGGGGCAA GTGGATGCG GTCTTCAAGG GCTTTTCCGA 300
 35 CTGTTTGTCT AAGCTGGGCG ACACATGGCC AACTACCCGC AGCTGGGAC GACAAGACGA 360
 ACATCAAGAC CGTGTGCACA TACTGGGAGG ATTTCCACAG CTGCACGGTC ACAGCCCTTA 420
 CGGATTGCCA GGAAGGGGCG AAAGATATGT GGGATAAACT GAGAAAAGAA TCCAAAAACC 480
 40 TCAACATCCA AGGCAGCTTA TTGAACTCT GCGGCAGCG CAACGGGGCG GCGGGTCCC 540
 TGCTCCCGGC GTTCCCGGTG CTCCTGGTGT CTCTCTCGGC AGCTTTAGCG ACCTGGCTTT 600
 45 CCTTCTGAGC GTGGGGCCAG CTCCCCCGCG GCGCCACCC AACTCACTC CATGCTCCCG 660
 GAAATCGAGA GGAAGATCCA TTAGTTCTTT GGGGACCTG TGATTCTCTG TGATGCTGAA 720
 AACACTCATA TAGGATTGTG GGAATCCTG ATTCTCTTTT TTATTTCGTT TGATTCTTG 780
 50 TGTTTTATTT GCCAAATGTT ACCAATCAGT GAGCAAGCAA GCACAGCCAA AATCGGACCT 840
 CAGCTTTAGT CCGTCTTCAC ACACAAATAA GAAAACGGCA AACCACCCCC ATTTTTTAAT 900
 55 TTTATTATTA TTAATTTTTT TTGTTGGCAA AAGAATCTCA GGAACGGCCC TGGGCACCTA 960
 CTATATTAAT CATGCTAGTA ACATGAAAAA TGATGGGCTC CTCCTAATAG GAAGGCGAGG 1020
 AGAGGAGAAG GCCAGGGGAA TGAATTCAAG AGAGATGTCC ACGGACGAAA CATACGGTGA 1080
 60

455

ATAATTCAGS CTCACGTGCT TCTTCCACAG TATCTGTGTT TGATCATTTT CACTGCACAT 1140
 TTCTCTTCAA GAAAACGCAA AGGACAGACT GTTGGCTTTG TTTTGGAGG ATAGGAGGGA 1200
 5 GAGAGGGAAG GGGCTCAGGA AATCTCTGGG JTAAGAGTAA AGGCTTCCAG AACACATGCT 1260
 GCTATGCTCA CTBAGGGGTT AGCTTTATCT GCTGTGTGTT ATGCATCCCT CCAAGTTCAC 1320
 TGCTTTTATT TTCTCTCTC CCTCTGTGTT TAGCTGTAC ACACACAGTA ATACCTGAAT 1380
 10 ATCCAACGAT ATAGATCACA AGGGGGGGAT GTTAAATGTT AATCTAAAT ATAGCTAAAA 1440
 AAACATTTTG ACATAAAGA GCCTTGATTT TAAAAAAGAG AGAGAGAGAG ATCTAATTTA 1500
 15 AAAAGTCTAT TATAAATTAA ATTCAGCAAA AAAACATTTG CTACAAACTA TAGAGAASTA 1560
 TAAATAAAA GTTATGTGTT GAAAAAAGAG AAAAAAAGW CTGACCGCA AGGGAAT 1617

20

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:
 25 (A) LENGTH: 1974 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GAATTGAGCA CGAGGCTGAG GGAAGTGCAG CGCAACAGAG TATCTGAGG CGCCAGGTTG 60
 CGTAGGTGCG GCAGCAGGAG TTTTCCCGGC AGCCAGGAGG TCCTGAGCAG CATGCCCGG 120
 35 AGGAGGAGCT TCCTGCGGC CCGCTCTGG CTCTGGAGCA TCCTCTGTG CTGCTGGCA 180
 CTGCGGAGGG AGGCGCGGCC GCGCAGGAG GAGAGCCTGT ACCTATGGAT CAGTGTCTAC 240
 40 CAGCAAGAG TACTCATAGG ATTTGAAGAA GATATCTGA TTGTTTCAGA GGGGAAATG 300
 GCACCTTTTA CACATCATTT CAGAAAAGCG CAACAGAGAA TCCAGATAT TCCTGTCAAT 360
 ATCCATGCA TGAATTTTAC CTGCAAGCT GCAAGGCAGG CAGAATACCT CTATGAATC 420
 45 CTGTCTGTGC GTTCTCTGGA TAAAGGCATC ATGCAAGATC CAACCGTCAA TGTCTCTG 480
 CTGGGAACAG TGCCTCACA GGCATCACTT GTTAAAGTTG GTTCCCATG TCTTGGAAAA 540
 50 CAGGATGCGG TGCAGCATT TGAAGTGGAT GTGATTGTTA TGAATTTGA AGGCAACACC 600
 ATTCTTCAAA CAGCTCAAAA TCTATCTTC TTAAAGATAT CTCAACAAGC TGAATGCCCA 660

60 GCAAACTGCT CAACACCTG TTCTAATGCA GAGAGCTTTT TATAAAGAG AAAATGATTT 720

TSCCCTCCAG GACTAGAGGG AGAGCAGTGT GAAATCAGCA AATGCCACACA ACCCTSTCGA 960
 AATGGAGGTA AATGCATTGG TAAAAGCAAA TGTAAGTKTT CCAAAGGTTA CCAGGGAGAC 1020
 5 CTCTGTTCAA AGCCTGTCTG CAGAGCTGGC TGTGGTGCAC ATGGAACCTG CCATGAACCC 1080
 AACAAATGCC AATGTCAAGA AAGTTGGCAT GGAAGACACT GCAATAAAAG GTACGAAGCC 1140
 10 AGCCTCATAA ATGCCCTGAG GGCAGCAGGC GCCCAGCTCA GGCAGCACAC GCCTTCACTT 1200
 AAAAAAGGCC AGGAGCGGCG GATCCACCT GAATCAATT ACATCTGGTG AACTCCGACA 1260
 TCTGAAACGT TTTAAGTTAC ACCAAGTTCA TAGCCTTTGT TAACCTTTCA TGTSTGAAT 1320
 15 GTTCAAATAA TGTTCATTAC ACTTAAGAAT ACTGCCCTGA ATTTATTAG CTTCATTATA 1380
 AATCACTGAG CTGATATTTA CTCTTCTTTT TAAGTTTCT AAGTACGCT GTAGCATGAT 1440
 20 GGTATAGATT TTCTGTTC AGTGCTTTGG GACAGATTT ATATTATSTC AATTGATCAG 1500
 GTTAAATTT TCAGTGTGTA GTTGGCAGAT ATTTTCAAAA TTACAATGCA TTTATGGTGT 1560
 CTGGGGCCAG GGAACATCA GAAAGSTTAA ATTGGGCAAA AATGCGTAAG TCACAAGAAT 1620
 25 TTGGATGCTG CAGTTAATST TGAAGTTACA GCATTTEAGA TTTTATTSTC AGATATTTAG 1680
 ATSTTTGTTA CATTTTAAA AATTGCTCTT AATTTTAAA CTCTCAATAC AATATATTTT 1740
 30 GACCTTACCA TTATTCCAGA GATTCACTAT TAAAAAATA AAAATTACAC TGTGGTAGTG 1800
 GCATTTAAAC AATATAATAT ATTCTAAACA CAATGAAATA GGAATATAA TGTATGAACT 1860
 TTTTGCATTG GCTTGAAGCA ATATAATATA TTGTAAACAA AACACAGCTC TTACCTAATA 1920
 35 AACATTTTAT ACTGTTTGTA TGTATAAAAT AAAGGTGCTG CTTTAGTTTT CTGA 1974

40

(2) INFORMATION FOR SEQ ID NO: 204:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1057 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

55

CGGCCTTCCG GGGCAACCGT TGTCCCAAC NCGGAAAGG GTCCTGGAGN CGGGAAC TAG 60
 GAGCCTCGGA AGTCCAAGG CAGAGCGGCC TTTGCTAATA AGCCAATCAG AACSTGAGAC 120
 GCTCCGGTGG GNCGGTGCCG TCGAGCGCGG GGTGGAGTCT GGGTGAATTG GCTGGCGGGA 180
 TCAAGTGAG CTGCTTCAGG CTGAGGTGGC AGATASTGAG CGCTGGTGGC GGASTTAAAG 240
 TYAAGCAGG AGAGTAATWA TGAATAGCGC AGCGGGATTC TCACACCTAG ACCGTCCGGA 300

60

	GGCGGTTTCTC AAGTTAGGGG AGAGTTTTCGA GAAGCAGCGG CGGTGCGGTT CCACACTGTG	360
	CGCTATGACT TCAAACCTGC TTCTATTGAC ACTTCTTCTG AAGGATACCT TGAGKTTGGC	420
5	GAAGKTGAAC AGKTBACCAT WACTCTGECM AATATAGAAA GTTGAAGGAA GCAGTAAAAT	480
	TCAGTATCGT AAAGAACAAC AGCAACAACA ATGTGSAATT CASCACGAC TCCCAATCTT	540
	GTAAAACATT CTCCATCTGA AGATAAGATG TCCCCAGCAT CTCCAATAGA TGATATCGAA	600
10	AGAGAACTGA AGGCAGAGGC TAGTCTAATG GACCACATGA GTAGTTGTGA TAGTTTCATCA	660
	GATTCCAAJA GTTCTATCATC TTAAAGTAGT GAGGATAGTT CTAGTGACTC AGAAGATGAA	720
15	GATTSCAAAT CCTCTATTC TGATACAGGG NAATTGTGTC TCAGGACATC CTACCATGAC	780
	ACAGTACAGG ATTCTGTATA TAGATGCCAG TCATAATAGA TTTCGAGACA ACAGTGGCCT	840
	TCTGATGAAT ACTTTAAGAA ATGATTTGCA GCTGAGTGAA TCAGGAAGTG ACASTGATGA	900
20	CTGAAGAAAT ATTGTAGTAT AAATAAAAAAT TTATACAGCA TGTATAATTT ATTTTGTATT	960
	AACAATAAAA ATTCTTANGA CTGAGGGAAA TATCTCTTAA CTTTTCATCA TAAAAGAAAT	1020
25	TAAATTTGAT TCAGAAAAAA AAAAAAAAAA AACTCGA	1057

30 (2) INFORMATION FOR SEQ ID NO. 205:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 721 base pairs

(F) TYPE: nucleic acid

(C) STRANDEDNESS: double

(I) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

40	GAATTCGCA CGAGTCATCC CTCCTCCCTCT TTCACTCCCT TACTCTTACT CTGTTTTTTG	60
	TGCTCCAGAC AGACAGACCC TACCTCTTTT GCTTCTTTT TTTTTTTTTG TTTTGAGATG	120
	GAGTGTGCT CTCTTTGCCC AGGCTGGAGT GCAGTGGGGC AATTCTGGCT CACCACAACC	180
45	TCTGCTTCC GCGTTCAAGC AATTCTCTTG CCTCAGCTC CCGAGAAGCT GGGGATTACA	240
	GGCATGCGCC ACCAGACCCA GCTHAATTTT ATATTTTATG TAGAGATGGT GTTTCTGCAT	300
50	GTTGGTCAGG CTGCTCTCAA ACTCCCAACC TCACGTGATN CCGCTTGCTT TGGCTTCCCC	360
	AAAGTCTGG GATTACAGGC CTGAGGCACT GCGCCAGCT TTTTTCCTC CTTTATACTC	420

60) BEANSTOCKS GENEALOGICAL RECORDS OF THE STATE OF MASSACHUSETTS 1850-1860 1870-1880 1890-1900 1910-1920 1930-1940 1950-1960 1970-1980 1990-2000 2010-2020 2030-2040 2050-2060 2070-2080 2090-2100 2110-2120 2130-2140 2150-2160 2170-2180 2190-2200 2210-2220 2230-2240 2250-2260 2270-2280 2290-2300 2310-2320 2330-2340 2350-2360 2370-2380 2390-2400 2410-2420 2430-2440 2450-2460 2470-2480 2490-2500 2510-2520 2530-2540 2550-2560 2570-2580 2590-2600 2610-2620 2630-2640 2650-2660 2670-2680 2690-2700 2710-2720 2730-2740 2750-2760 2770-2780 2790-2800 2810-2820 2830-2840 2850-2860 2870-2880 2890-2900 2910-2920 2930-2940 2950-2960 2970-2980 2990-3000 3010-3020 3030-3040 3050-3060 3070-3080 3090-3100 3110-3120 3130-3140 3150-3160 3170-3180 3190-3200 3210-3220 3230-3240 3250-3260 3270-3280 3290-3300 3310-3320 3330-3340 3350-3360 3370-3380 3390-3400 3410-3420 3430-3440 3450-3460 3470-3480 3490-3500 3510-3520 3530-3540 3550-3560 3570-3580 3590-3600 3610-3620 3630-3640 3650-3660 3670-3680 3690-3700 3710-3720 3730-3740 3750-3760 3770-3780 3790-3800 3810-3820 3830-3840 3850-3860 3870-3880 3890-3900 3910-3920 3930-3940 3950-3960 3970-3980 3990-4000 4010-4020 4030-4040 4050-4060 4070-4080 4090-4100 4110-4120 4130-4140 4150-4160 4170-4180 4190-4200 4210-4220 4230-4240 4250-4260 4270-4280 4290-4300 4310-4320 4330-4340 4350-4360 4370-4380 4390-4400 4410-4420 4430-4440 4450-4460 4470-4480 4490-4500 4510-4520 4530-4540 4550-4560 4570-4580 4590-4600 4610-4620 4630-4640 4650-4660 4670-4680 4690-4700 4710-4720 4730-4740 4750-4760 4770-4780 4790-4800 4810-4820 4830-4840 4850-4860 4870-4880 4890-4900 4910-4920 4930-4940 4950-4960 4970-4980 4990-5000 5010-5020 5030-5040 5050-5060 5070-5080 5090-5100 5110-5120 5130-5140 5150-5160 5170-5180 5190-5200 5210-5220 5230-5240 5250-5260 5270-5280 5290-5300 5310-5320 5330-5340 5350-5360 5370-5380 5390-5400 5410-5420 5430-5440 5450-5460 5470-5480 5490-5500 5510-5520 5530-5540 5550-5560 5570-5580 5590-5600 5610-5620 5630-5640 5650-5660 5670-5680 5690-5700 5710-5720 5730-5740 5750-5760 5770-5780 5790-5800 5810-5820 5830-5840 5850-5860 5870-5880 5890-5900 5910-5920 5930-5940 5950-5960 5970-5980 5990-6000 6010-6020 6030-6040 6050-6060 6070-6080 6090-6100 6110-6120 6130-6140 6150-6160 6170-6180 6190-6200 6210-6220 6230-6240 6250-6260 6270-6280 6290-6300 6310-6320 6330-6340 6350-6360 6370-6380 6390-6400 6410-6420 6430-6440 6450-6460 6470-6480 6490-6500 6510-6520 6530-6540 6550-6560 6570-6580 6590-6600 6610-6620 6630-6640 6650-6660 6670-6680 6690-6700 6710-6720 6730-6740 6750-6760 6770-6780 6790-6800 6810-6820 6830-6840 6850-6860 6870-6880 6890-6900 6910-6920 6930-6940 6950-6960 6970-6980 6990-7000 7010-7020 7030-7040 7050-7060 7070-7080 7090-7100 7110-7120 7130-7140 7150-7160 7170-7180 7190-7200

TACACTCAGC CTGGGCAATA GAGGGACATG TTGTCTCTAA AAAAAAAAAA AAAAAACTCG 720

A 721

5

(2) INFORMATION FOR SEQ ID NO: 206:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

CCACCATTTA TCCAAC TGAA GAGGAGTTAC AGGCAGTTCA GAAAATTGTT TCTATTACTG 60

20

AACGTCTTTT AAAACTCGTT TCAGACAGTT TGTCTGAACA TGAGAAGAAC AAGAACAAAG 120

AGGGAGATGA TAAGAAAGAG GGAGGTAAAG ACAGAGCTTT GAAAGGACTT TTCCGAGTGG 180

25

GAGTATTGGC AAAAGGATTA CTTCTCCGAG GAGATAGAAA TGTCAACCTT GTTTTGCTGT 240

GCTCAGAGAA AGTTTCAAG ACATTATTAA GCCGTATTGC AGAAAACCTA CCCAAACAGC 300

30

TTGCTCTTAT AAGCCTGAG AAGTATGACA TAAATGTGC TGTATCTGAA GCGGCAATAA 360

TTTTGAATTC ATGTCTGGAA CCCAAATGC AAGTCACTAT CACACTGACA TCTCCAATTA 420

TTGAGAGAAG GAACCTGAGG GAACGAGATG TAACCTCGGG TATGGTGAAA GACCCACCGG 480

35

ACGTCTTGGG CAGGCAAAAA TGCCTTGAGG CTCTGGCTGC TCTACGCCAC GCTAAGTCGT 540

TCCAGGCTAG ACCTAATGGT CTGCAGTCCT GTGTGATTAT CATACGCATT CTTGAGAGCC 600

40

TCTGTGAGCG AGTTCCAAC TGGTCTGATT TTCCAAGCTG GGCTATGGAG TTACTAGTAG 660

AGAAAGCAAT CAGCAGTGCT TCTAGCCCTC AGAGCCCTGG GGATGCACTG AGAAGAGTTT 720

TTGAATGCAT TTCTTCAGGG ATTATTCTTA AAGGTAGTCC TGGACTTCTG GATCCTTGTG 780

45

AAAAGGATCC CTTTGATACC TTGCAACAA TGAATGACCA GCAGCGTGAA GACATCACAT 840

CCAGTGACAA GTTTGCATG AGACTCCTTG CATTCCGCCA GATACACAAA GTTCTAGGCA 900

50

TGGATCCATT ACCGCAAAATG AGCCAACGTT TTAACATCCA CAACAACAGG AAACGAAGAA 960

GAGATAGTGA TGGAGTTGAT GGATTTGAAG CTGAGGGGAA AAAAGACAAA AAAGATTATG 1020

ATAACTTTTA AAAAGTGTCT GTAAATCTTC AGTGTAAAAA AAACAGATGC CCATTTGTTG 1080

55

GCTGTTTTTC ATTACATAATA ATGTCTACAT TGAAAAATTT ATCAAGAATT TAAAGGATTT 1140

CATGGAAGAA CCAAGTTTTT CTATGATATT AAAAAATGTA CAGTGTTAGG TATTATTTGA 1200

ATGGAAAGAC ACCCAAAAAA AAAAATGTGC TCCGACTAGG GGGAAAACAG TAGTTCCGAT 1260

60

TTTTTCOCAT TATTTTATT TTATTTTCTG GTTGCCCTAG CTTCOCOCOC TATTTTGTG 1320
 TCTTTTATTA ACTAGTCAT TGTCTTATTA AATCTTCACT GTATTTAATG CAGGATGTGT 1380
 5 GCTTCAGTTG CTCTGTSTAT TTTGATATTT TAATTTAGAG GTTTTGTGTTG CTTTTTGACA 1440
 CTAGTTGTAA GTTACTTTGT TATAGATSGT ATCCTTTACC CTTTCTTAAT ATTTTACAGC 1500
 AGTACGTTTT TTTGTAACGT GAGACTGCAG AGTTTGTMTT TCTATATGTG AAGGATTACA 1560
 10 ACACAAAAAG TTATCCTGCC ATTCGAGTGC TCAGAACTGA ATGTTTCTGT AGATCTTGTG 1620
 GCATTTGTCT CTAGTGTGAT ATATAAAGST GTAATTAAGA CAGAGTTCTG TTAATCTAAT 1680
 15 CAAGTTTGCT GTTAGTTCTG CATTAGTACT ATAAAAGCTA ATATATACTA TATGGTCTTG 1740
 CAACAGTTTT AAAGCCTCTG CATAATTGAT AATAAAAATG CATGACATTC TTGTTTTTAA 1800
 TAGACTTTTA AAATCATAAT TTTAGGTTTA ACACGTAGAT CTTTGTACAG TTGACTTTTT 1860
 20 GACATAGCAA GGCCAAAAAT AACTTTCTGA ATATTTTTTT CTTGTGTATA AGTGGAAAGG 1920
 GCATTTTCA CATATAAGTG GCTAACCAA TATTTTCAA AGAACTTCAT CATTGTACAA 1980
 25 CTAACAACAG TAACTAGGCC TTAATTATGG TGACAGTTCC TTATTGGTGT GTGTGAGATT 2040
 ACTCTAGCAA CTATTACAGT ATAACACAGA TGATCTTCTC CACACACCCC ATCACCAG 2100
 TAATTTACAG TTCTGTAAAC AGTGAGGTG ATAAAGTATT ACTGATAAAA AATTATCTAA 2160
 30 GGAAAAAAC AGAAATTAT TTGGTGTGGC CATCTTACCT GCTTATGTCT CCTACACAAA 2220
 GCTAAATATT CTAGCAGTGA TGTAATGAAA AATTACMTCT TACTGTTGAT ATATGTATGC 2280
 35 TCTGGTACAC AGATGTCAAT TTGTTGTAC AGCACTACAG TGAAATACAC AAAAAATGAA 2340
 ATTCATATAA TGACTTAAAT GTATTATATG TTAGAATTGA CAACATAAAC TACTTTTGCT 2400
 40 TTGAAATGAT GTATGCTTCA GTAAATCAT ATTCAAATTT AAAAAAAAAA AAAAAAAAAA 2460
 CTGGA 2465

45

(2) INFORMATION FOR SEQ ID NO: 207:

(1) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 1480 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

GGTGTTTCTTCTTAAATTTTATTA AATCTTCACT GTATTTAATG CAGGATGTGT
 1480

5 GCTGGCTGAA CGAGAGCAGG AAGAAGCCAT TGCTCAGTTC CCATATGTGG AATTCACCGG 240
 GAGAGATAGC ATCAGCTGTC TCAGCTGCCA GGGGACAGGC TACATTCCAA CAGAGCAAGT 300
 AATGAGTTG GTGGTTTGA TCCACACAG TGATCAGACA TTGGGCCCTC AGCGAACTAA 360
 GCAATATGTC CTCTGTCCA TCTGTCTTG TCTCCTGGCA TCTGGTTTGG TGTTTTCTT 420
 10 CCTGTTTTCG CATTGAGTCC TTGTGGATCA TGACGGCATT AAAATGGTGA AATCAGATT 480
 TAATAAGCAA GACTCCCTTG TAATTCTCAC CATCATGGCC ACCCTGAAAA TIAGGAACTC 540
 CAACTTCTAC ACGGTGGCAG TGACCAGCCT GTCCAGCCAG ATTGAGTACA TGAACACAGT 600
 15 GGTGAATTTT ACCGGGAAGG CCGAGATGGG AGGACCGTTT TCCTATGTGT ACTTCTTCTG 660
 CACCGTACCT GAGATCCTGG TCCACAACAT AGTGATCTTC ATGCGAACTT CAGTGAAGAT 720
 20 TTCATACATT GGCTCATGA CCCAGAGCTC CTTCGAGACA CATCACTATG TGGATTGTGG 780
 AGGAAATTC ACAGCTATTT AACAACTGCT ATTGGTTCTT CCACACAGCG CCTGTAGAAG 840
 AGAGCACAGC ATATGTTCCC AAGGCTGAG TTCTGGACCT ACCCCACGT GGTSTAAGCA 900
 25 GAGGAGCAAT TGGTCACTT AACTCCAGC AAACATCCTC CTGCCACTTA GGAGGAAACA 960
 CCTCCGTATG GTACCATTTA TGTTTCTCAG AACCAGCASA ATCAGTGCCCT AGCCTGTGCC 1020
 30 CAGCAAAATG TTGGCACTCA ATAAAGATTT GCAGAATTTA ATACAGATCT TTTGAGCTGT 1080
 TCTTAGGGCA TTATAAATG AAATCATAAC GTGTTCTAG GTTATCAAAC CATGGAGTGA 1140
 TGTGGAGCTA GGATTGTGAG TACCTGCAG GCCATTATCA GTCCCTCATC TGTGCAGAAG 1200
 35 TCGCAGCAGA GAGGAGCTAT CCAATACCT AAGAGAAAAC AGACCTAGTC AGGATATGAA 1260
 TTTGTTTCAG CTGTTCCCAA AGGCCTGGGA GCTTTTGA AAGAAAGAAA AAAGTGTGTT 1320
 40 GGCTTTTTTT TTTTATGAA AGTTAGAATT GTTTTACCA AGAGTCTATG TGCGGCTTGA 1380
 TTCACCTTC ATCCATGGC TGAACATGG ATTGGGGATT TGATAGAAAA ATAAACCTG 1440
 45 CTTTTCATTC AAAAAAAAAA AAAAAWAAA AAAAATCGA 1480

50 (2) INFORMATION FOR SEQ ID NO: 208:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 872 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

60 CAGTATTTCC CTCAGTACTG TAAGCAAAAG TGGTATGTTT TTCTTTCTTT ATGTCTACTC 60

	GTGCTCTGTGT GGGCTTCTGG TGTAACCTCT TCTTCCTAGC CATTTCAGTCT CTCTAGTCAAC	120
	CTCCCTAGTA GCTAGTGGTC TCTAAGTTTT TATTTAATTA GAACAACTCC ATTTCCATTT	180
5	CAAGGTAGGT CAATGGGGGG AAAAGCCTCA TGATTTTAAAC TGAAGTTAAC AACACAGCTT	240
	TTAAATGAA AACTCATACT CCAACTTCTA AAGTATATTT GAGCTGATTT GTTTCCAAAA	300
	CAAGATATG CTGTACCTAA AACTGCTAAA ACAAAAATAT AAAGACAAGG ACTAGGTGAT	360
10	TAACGGGAGA GAAAAATCAT YTCTTTTCCA GGAAACCTTT CCTAAAATAA GCAAAACTTG	420
	ANTCTATGCT TCATGGAAAC TGACACAAAG AAAAGAAACT GATGGATTGC ACAGGCCTTG	480
15	TTATAGAAAT AGATCTATAA AAAGATCTGT CCACAGGAAA TATACACCTT CTCCTGGTTC	540
	TGAACCTCAA TGGGGATTTG TCACCTAGGT CTCATCTAT AGGAATACCT TCACATACCT	600
	ATCTATTGAT GCALATATTC TGAAAACAGG TACATACAAA ATTACAACAA AGGAAAAAAA	660
20	TTCTATTGAA CACTTAAAAA TAGAAACAGG CCAGGCACGG TGCTCATGC TGTAAATCCA	720
	ACAATTTGGG AGGCTGAGGC TGGTGGATCA CCTGAGGTCA GGA CTGTGAG ACCAGCTTGG	780
25	CCAACAIGGT GAAACCCCGT CACTACTAAA AATACAAAAA AAATTAGCCT GTGTGGTGGC	840
	ACACTCCTAC AATCCNGCCT GACTCGGAA AN	872

30

(2) INFORMATION FOR SEQ ID NO: 209:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1779 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

	ANTTGGCAAG ACTGCACAAA ATTACAGTGC TAATGTATAT G3TTGCAGTT CACATAAAGA	60
	CAAAAGCATC TGTTAATGAAA TTAGTAGTAA TATTGGGTGG TTGATTGTGT CTTAGCAGAC	120
45	TTGGCTTCAT WTTGGTCTTG AGATAAAATG GCCAGCATAA ATGCTGTTTA TATTACGTT	180
	TTCTTAGGTG TGTGTGTGCA GGCACAGCA GCATGCCCTT GGTGTAGTCA GTGCCGAAAS	240
50	GGGTCTGTTC CTTCTTAGC CTGCCTGCAG GGATGCTCTC CTTTAAAGC AGGTGTGTG	300
	GAGCATTGAG TACACTAAG GTAAGCTAAA CCATCAACAT CTCGTGTTT TTAAGATGTT	360

(6) TAACTTCTGTTT TAATTAAGT TACNNTTTTT AAATCAATT GGTATAACT TTATATGAAT 500

5 TTTGTAAAT ATAGTTCCTA GTGACATAGA AACGATGCGT AGTTTTGATT TACTAATTAC 660
 AAATGTTGAG GCCTAATTCT GAAAGTCCTC ATATTTAAAG GCTAGACAAC GTAATGAAAT 720
 TTTTAACTAT TTGTATGTCA TTTTBAAGT GTACTGCTTT ATGGTAAAAG TGTTTTTCAT 780
 TTGTTTCATTG TTTTCATTAT TTGTATCAT GTTGTCTTTC AATACAGGCA TAAACCTTCC 840
 10 ACTCTTGAAC AAAGCAGCTG CTTTTTAAAA GCGGTAATTG CTTCTTTACC TTTTATTTCT 900
 TTGTGAAATG AAGCTTTTCT TTAAGAATGT GACTTTAAAG TGTGTCTAT TGCATAAAAC 960
 AGTTGACACT CACTTATTGT AAAGTBAAGA TTGTTCTACT GCATGTGAAG TGGACCATGC 1020
 15 AGATTTCTGT ATGTTCTCAG TATGCATCAC TAGATAATAA AGTCTTTTGT GAACAAGGCA 1080
 TTTGTAGCCA TTTTAAAAAG TTTTGTCTTT CAGTGCTGGT AAGTCAGGTA AACCATAAAT 1140
 20 AGTTAAAAGC AACCTTTTGT TTTTTCCTG AAAGTTTTTA ATTGAAAGTA TTATTAGTTA 1200
 AAGATGTAAA CCTAGCCAAA ATTACCAGTT TATTAATAAT TAGGATCCTA ATTATTTCAA 1260
 AAAATCCTAC AAATATTGTC AGTTTTCAGT GTAGTGAGAT TATTCCTGTA GGTATGCGG 1320
 25 TATAATTCAG GATTAACTA ATTTTCTGTC TATTTCTCA CTTTTCCTTT TGATGGTGCG 1380
 GAAAGAGAAA AAGGAAAACG GGGCACAGGC CATTCGACGC CTTCTCCAAG GGGTGTGATT 1440
 30 TGCTGAGACA CCAGCTTCAC CTTTMTAACA AGGCACCTAA TTACAACAAG CATGCACATT 1500
 TTGGTGCAAT CAAGAATGGA AAATCAGAAT AGCAGCATTG ATTCTTCTGG TGCAGCTCAG 1560
 TGGAAGATGA TGACAACCAG AAGACATGAG CTAAGGGTAA GGGACTGTTT TGAAGAACCT 1620
 35 TTCCATTTAG TGAATCAAGAT ATGGAAGCTG ATTTCTGAAA ATGCTCAGTG TGTACTCTAA 1680
 TTATTTATGG TACCATTTGA ATTGTAACCT GCATTTTAGC AGTGCATGTT TCTAATTGAC 1740
 40 TTACTGGGAA ACTGAATAAA ATATGCCTCT TATTATCAA 1779

45 (2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2110 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

55 GCGGCCGCTG CAGCCCGGAG CTGAGCTAGC GGTCCGAGCC GAGCCGTCCG AGCCGGGGAA 60
 GCCGCGCGGT GCTGCCGCTC GTGGCGGCCA GAGGAGAGGA GAGGCAGCAG CATGGCGAGT 120
 60 GTCCTGTCCC GACGCCTTGG AAAGCGGTCC CTCCTGGGAG CCCGGGTGTT GGGACCCAGT 180

	GGCTGGGAGG GGCTGGGCT GGGGCACTCT CGGAGCCACT GGTAGAAGG GGGCTGCCC	240
	AGCTTTTAC CAGCTGTAT GACAGCCCTT GCGAGSACA GCGCAAGGAA GTCTTAAGG	300
5	CTCCAGTAC CTGGGCTTT CAGCAGGTGG CTTTMAAGC TGCGCAGAAG GTTTATGTGT	360
	GTAAGGGGG TCAAGAGTGC ACAGGATTGG TGGGCGAGCA CAGCTGATG GAGGTGAGG	420
	TGAGGTGTG GGTGTGTAG CACAAGCTGT AGTCTGTCTG CAGGTGTAG GAGGTGTCTG	480
10	TGCGAGGCT GCGAGGCTCC TGTCCCGAG GAGCAGCTCT GGAGCCCGCA GCGCAGCCCT	540
	TGGCTACAG GCGGTGTCT AGGACATGG ATGTCCAAA GAGGAAGTG GAGGCATGA	600
15	AATGATGAG ATGATGCGG CAGTGTCT GAGTCTCTG TCTGCAAGC CTGTGTGTA	660
	GAGTCTCTC GCGAGCGAG CCAACTTCT TGTTCCTCT GCGGCTGAG ACCCATGGAA	720
	CGAGGTGCT GACATCTCG ACAGCGGAG CAGCACTACC ACCGTCACT GAGTGTGAG	780
20	CAGTGTGTC TCGAGCTCT CCGGCGGCA CCGGCGGAG ACCCGCAAT ATTGCGGGA	840
	TGCTTTGCT TCTCCGAAA CTGATCATG CTGTGAGAG GATCTGAGC CTTCTCTCT	900
25	CGAGGAGCA GCTCGAGAA AAAGAAAGAA CTCTGTGAG GTGATGTACA ATGCTGTGT	960
	GCGAAATCT GCGAAATCT TGGCTCTAT TGTGGGATC AAACGACAG TCAAGGCTCT	1020
	CGATCTGCG GAGCAGTGG ACTCTGATC GTTCAAGCG GAGGAGGATT TGTACTAGC	1080
30	AGAGGTGTG GTGAGGTAG AATCTGTCT TGTGTCTCT GTGTGTCTG CAGACCCCA	1140
	GTCTGTGCA CTCTCAGCT CAGGCGAGT CCGACCCCA GATGAGTGG CTGTCTCTG	1200
35	TCTGTCTTC CAGCAGCTCT CACAAAGCC CAGTCTCTG GTCCAGACA TCTGTGCGCG	1260
	GAGTCTCTC TGGCTGTAG GGTCTCTAG AAGTCAGCT CTGGTCTCT CTGGCAGAT	1320
	CAGGAGATC ATGCATACA GGTCTGTCA TCTTCCAGA TCGAGTCTC ACCACATC	1380
40	TACACAGTG TCAGCTGGG TGTGTGCGG TCGCGCGCT GTGTCTCTC TCGGTCTCG	1440
	AGCGGTCTG TATGCTGAG CAAAGCTCA CAGGCGAGC CCGGATGA AATCTCATCT	1500
45	GATGCTACT TGTGAGCG GCGGCGAGG TGGTGTGAG AAAGCCGAG GCGAGGTAA	1560
	GAGTGTGCG AAGTGTAGG CATGTAAG CCGGAGCAT GTGTGAGCT TCGCGGTGG	1620
	AAGAGGTCT GCGAGGCTT TGTGAGTGA TGTGTCTCT AGGTCTACT CTGTCTCTG	1680
50	CCTGTCTCG AGCAATGAC AAGAGGTAG TGTGTACCA GCGCTCAGC AAAACCGAA	1740
	GAGAAAGAA GGAAGACAG AGTTTGGCT CTGTGTGCTA AGGTGTAAC CTAAAGCAA	1800

AAAAGAAGGA ACAGCTCGTT CTGCTTCCTG CTGAGTCGGT GAATTCTTTG CTTTCTAAAC 2040
TCTTCCAGAA AGGACTGTGA GCAAGATGAA TTTACTTTTC TTAAAAAAA AAAAAAAAAA 2100
5 AAAAACTGA 2110

10 (2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 938 base pairs
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

20 GGCACAGGAA AAAAAAGAAA AAAGAAAAAA GAAAAAGTT TTTGTACCCA CAGATTAGCA 60
TTTCTTNGAT GTTTGAAAAA AGTTTAAGCT ATGTCCTAAT TTAAAAATGA GCACAAACTA 120
CTTAACAGAT GTGTGTTCCC TCTTCTCTTA CTAAATTAT CTTTATTTTC ACCATCACCT 180
25 CCCAGTGTCG AACACCTGAN CTGTGTGTTT TGTGGTTGGA TCCTGGGTTG CCAAGTTCCT 240
ATTTGGTTCAG TCCCTGGCCT GTGGGGCGGT CTCAGGAAGT GGCATGCTCT TCAMGRAGGA 300
30 TCGTTCATYT CCAGTATAAC CAWTTTGTTA ATAATAGTTG ATAATTCCCA GCTTTTACCA 360
GATGAPTTTT GACTTATTTT TCCCTCTTTG ACCTGTTCOA AGCTAACATA TCTCGGTCAG 420
TTCCGAGAGG GTGGGGGATT TGACAATGTG AGGAGGAGTG GGGTTAGAAT GGGTTTGCTT 480
35 ATCTGGGCAA GGAAAGAGTT CCTAGTCGAT TGGGCACAAT GACAAAATGA TTCCATGGAT 540
AGAATCGTCC CATGTTGCTG GAACACCTCA CGTGTGTGTA ACGCCTTAAA TTCCTGCCAT 600
40 CCGTTCTCTG ATTCCCCACC TCCCTGTAGT TTCCACAGGA TTTATCTCTC TGTACCCCCG 660
TCCTCCAACCT CTA CTCTGTC AGCCTCTCCT CCATCCCTTA CTTCCCTTCT AAATTCAGG 720
AGATGACCTC ACTTTGCAAA GCAAAATGGA GCCACCAAAT TGTAGCTCTC CTCGGTGGAA 780
45 ACTGCATCTG TGTCATCCC TGCACCTTCT TGCAGAAAGC CGCCCCCTCA GGCCAAGATG 840
AGTGCCCTGGC CCCCATGGGA GACTCAGACA CTTTGACCCC TTGTGACTTC AGCATCTCCC 900
50 TCTTTAAAGA TTCTCTCCCA ACAATCAGTC GTGCTCGA 938

55 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1551 base pairs
(B) TYPE: nucleic acid
60 (C) STRANDEDNESS: double

(E) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

5	AGGCTGGACT AAGCATAGAG AACCAAGAGA GAAAGAAAGA TTAAAGAGAC TGAGTAATAT	60
	TTTTTGACAG ATCATTTAAG AACTGAGTA ATTTTTTTTT TCTCCAAAAG GGCATGGGT	120
	TTTTTTTTGT TTTTTTTTTT CTCTATTTGG CACTTTCTAG GGATTGGTCT ATAAATTTTT	180
10	TCAAAGATCA TAGAATAAAT TTCTTTGTAG CAACTTCCTA TTTTAGTGTT TATTTAGGS	240
	GARCCCCARG TGTCCTTGCT GATACGCCAT TAGGGCCACT TCTCAGCCTC TGGTTACATC	300
15	ATAATGCTTT TTTTCTATC TTGCCAAAGT TTTCMGAAAA TTKAGTTTTT CTAAATTTAA	360
	AAAAATTTGGT TGTGGAGATG GATGCGGACC TCTTTATAAG CCTGAAAAT AAGTGATTTN	420
	TTTTAAGTGC TATTCTGCTA TAAACCTGAT TCTCACTTTT TTCTGTAGAC AACAGTTTTT	480
20	TATAATATAT CTATTTTCTG TGGACATTAT TTCCTTTTAA CCAATACTGA AATTCCATAG	540
	TGTAWACITT CTTCACATTT TCTTTGATTA ATACTTYCTT AAAATAGACA CTTCGATTGG	600
25	CACCAGCTGT CACCAATAAA GCTGCCCTGA ACATTGTCAA TCAATCCTGT TAACCAATTT	660
	GAGAATTTTT CTGGAATGCT TAGTTAGGGA TGAAATTGCT GGGTTATAGS TATGAGTATG	720
	CTTGATATAC TTTTCTCCAG AATGCTTACA CCTGTGTGTA CACCACATCT CCAGAGATAG	780
30	GGGAATCTTA TGTCCCTGCT AACTGCTCTC GTTATTTAAT TTTCTGACAT TTGCCGCCGC	840
	CGCCGCCCCC TGCCCCCAAC ACACACATGG TATAAAGTGG TAGTTTCTTG TTTTAAATG	900
35	AACTTTGAAG TGATTTGAAT TTGGGCATTT CTTTGTATCC TGAGTTATTT TGTTTTCCCG	960
	TTATGTGAAT ATCTTTTCC TATGCTTTAA CTACTTTTCT AATTTGTCCC TTTTTTNGGT	1020
	TATCAAATTC CAGGCCATTG TCTATTCCAT CGTCACTTT GGGTATTGGA AACATCTTTC	1080
40	CATTCTGTAG CCGTCTGTGT GAACATAAAT CTGTATTTTT ATGTAATCAG ATTTTCTCC	1140
	TTACGGTAT ATTTCTTGAA TTTTATTTAA GAAATCTTTT TCTATCCTGA GACCACAAAA	1200
45	ATGTCCCCAG CATTTTCTTC TGTTTCATAG TTTTCCCTTG TATGTTTAAAT CTTTAAAGGC	1260
	ATGTGTAGTT CATTTTATAT GGTGTGAAAT AGTTCTTATT CATTTATTTA ACACATATG	1320
	GTGGAGTGCC TGCTGATGGT AGTACTCTTC AGAGTACTTT GTATATATTT GTGAACACAT	1380
50	ATTCTTGCCC TGAAGCTTA TGTTGTCTTT CAAGGTAGAT CCACTACTGG TTCCACCTG	1440
	TTTTCTTTCA CCGTACGGAT GAATTCACCA ATTTTACACA TAGCACCAGT TAAGGAATAG	1500

60 (ii) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 997 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

10 AGAGAGTCCT CAACAGAACG TAATCATGCT GGCACCCCTAA TCTGATACCT CTAGCCCTCGA 60
 GAACTCAGAG AACATAAACT CCAGTTCTTT AAGCTACCGA CCTATGCTA TTGTTCTATTA 120
 TAGCCCCAAGC TAAATCAGGT GGAAAGGCG AGAATATTTTG AAAAATCTCA TTTCTACAAA 180
 15 AACAGAGTTG TTCTAAATGA AATGGCCAGA TATTCATCT TCTGATACCT AGTATTTAIG 240
 AAAGTTTCAT TAAACAGCA TGGCCAGCA CCCAGGCGTG CCGCTTCAG AACGGCAAG 300
 20 AAAAGCAAAT GATTTGAGGA ACAAAGAGT GGACACAGAG CTTCTCAGAA GATGGCTCGA 360
 TCTCTGAGA TGATCTCTG AGATCATCAA TTTCTGCGAC CTGATGTCCT ACTCCAAITG 420
 TAGTAGATAA GAGCAAAGAC ACTTCCTGAT CCGTGTGAAA ATGCTGGAGC CTGCTGTATG 480
 25 GAGAGGCTGA CACTGGGAGC AACAGAAGGC GGGACATTTA TTGCTGCAAG CCTCTCTGCA 540
 CCTGGGCCCT CTTCAGGCT TGTACCTTGC ACTCCGCAAG CCACTGTAGC AACTGGTAA 600
 30 CTGAAGTTAG GTATTGGAAG AGATAATTTG CCCCCAACA AGAATTAATT AAAAGAAAA 660
 GGAAACCACT AAATTCGCT TGACAAACCA GTTGTGTCAG TTGTAATTT TCAAAATTTG 720
 AAACTTCTG TTGCGCAGCA TATGATCTG TTACATTAGG GTCATCAAT CTAAGATAC 780
 35 ACAGCTAGGT TACCAAGCT CCAGTGCTCA AGAATGAAG AACCTCTGAG AGAGAGATCA 840
 GTTTCTAATA ACCTAACAT TTTCTTGGG TATTACMAA AAAAAAAA TTAGAATAAA 900
 40 ATGTCAGTGC CATGCAGGCA AGTACAGATA TGGAAATGA AGCTCTGTCT ACAAATCGAA 960
 GATTGTGTTG TTAATAAAAT TGATTGGGAT CACTCGA 997

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1496 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

50 GAATTCGGCA CGAGTGACCA CAGATATCTT TGGCTTTCAG CCTCAGCACA ATGCTGTCCA 160
 CTATGTTTTT TTAATCGAT TGACATCTCA TGAATCCACA AATTAACCG GTTCTCCATC 120
 60

	TTTTCCATCT TTGTGATAGC TTCATCAAGC ACCATGGAGG TCACTTCAGC ACTATCCGGA	180
	GGGGCTCAC GGACAGATCT GTGAATTTCG TTTTCCTTTT TCTTGATGTA CCGGATTGTC	240
5	GACTGGTTAA CATTGAGCTC ATGGCCAAIA GCACTSTAAC TCATGCTGA TTEGAGETIA	300
	TCCAACACGC GGATTTTCTC CGTAAGGSAM ATCAMEGTCT TCTTTCGCTT AGGAACACTG	360
	GGCARACTTT AARCACTAGC CTGGGGGGGC ATTTTAGAAA GCAAAACCAC CCACAAAAAG	420
10	CAGAAAAAAA AGTGTCACTA AACAGACTGN NGANAGGACT CTTTSTTTAC AGCAGAGGAG	480
	CTGCGACTAG AAGCGGGGGC TTCTCCCGAG TTCAACTTC AGTGGGAAC CTTACCTCCG	540
15	CGAATCCAA ATTTTCACGC TCTCCCATG CCCCGGAAAS AAACCCCCAG AACAGTACGS	600
	TCATGATGTA TTTTAGCGTT ACAATACAT TTTACCAAGT AATGGAATTT GGCATTACGA	660
	ATTATGATT AATGAAGTC ACCTGTATTT CATTAGATAT GTAATTTTAT TTAGGAGGT	720
20	TTATTATATT AAGGCGGGA GGCAGCGGCG AAGACTAGAA GTTCAGCAT GACCGGGTC	780
	CCGCGGGGTT CGGCTGCGA CGGAGGGGTT CAGGACGCC AGCGCGGAGG CATCGCGCGG	840
25	AAGTTCGTA GCGCACTAC GTASTACTCT CTGGCATGT GCAAACCGCT GTGGGGGCGC	900
	GGCTAGCTG CCGTGGGGC CCGCGGGGTT CTATGCTCT TCGTAGAGC TTTCCCTTG	960
	GAGCGGCTG CTGGGTCTT GTGAGTTTGA CCAGCGTGA GCGCGAGCAA CATEGAGAA	1020
30	TTGACTCCG AAGACTTCT TACGTGGAG GAGCAGGAG ACTACGTGC GTGCGGTGAG	1080
	CGATTCGGC TCGCGGAGA AGCGAATTC CCGCCGAC GCTCACGTG AGCGCGGTC	1140
35	TGCCCCCGG GCGGTCTGCC CTGTGGGCGA GGTGTTCAG CGCGGTCTT GTTCTCGAGC	1200
	GTGCGCTCC TCGCGGCTT CATGCTGCG CCGTCCGGC CGAGGCTGT GCGGTGCGG	1260
	GTTCTGTCT CCGCTGCTT TGGCAGCTC CGCGCGGCG CCTCTTTC ACGCGGGA	1320
40	CGGCACATGG AAGCGGCGC TTGTGCTAG GAGCGCTCT GGTACAGCC CAAAGACAA	1380
	CGCTGCTTCA GAAGTCGGG CGGCAGTTC AGCTTGGAA GTTTTTCCTA CCGCTGCTT	1440
45	GAGAGAGCTG CTGCGTACA ACGGTGCAA GATAGAGCTG TCGTCTCTC GCTGCTG	1496

50 (2) INFORMATION FOR SEQ ID NO: 215:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1398 base pairs

60 (3) INFORMATION FOR SEQ ID NO: 216:

CTGCCTTTGA CCCATCACAC CCGATTTTCCT CCTCTTTCCC TCTCCCGGCT GCCAAAAAAA 120
 AAAAAAAGG AAACGTTTAT CATGAATCAA CAGGGTTTCA GTCCTTATCA AAGAGAGATG 180
 5 TGGAAAGAGG TAAAGAAAAC ACCCTTTGTT CCCAACTCCA CTTTACCCAT ATTTTATGCA 240
 ACACAAACAC TGTCTTTTGG GGTCCCTTTC TTACAGATGG ACCTCTTGAG AAGAATTATC 300
 10 GTATTCACAG TTTTLAGCCC TCAGGTTACC AAGATAAATA TATGTATATA TAACCTTTAT 360
 TATTGCTATA TCTTTTGTGA TAATACATTC AGGTGGTGCT GGSTGATTTA TTATAATCTG 420
 AACCTAGGTA TATCCTTTGG TCTCCACAG TCATGTTGAG GTGGGCTCCC TGGTATGGTA 480
 15 AAAAGCCAGG TATAATGTAA CTTCACCCCA GCCTTTGTAC TAAGCTCTTG ATAGTGGATA 540
 TACTCTTTTA AGTTTAGCCC CAATATAGGG TAATGGAAAT TTCTTGCCCC CTGGSTTCCC 600
 20 CATTTTACT ATTAAGAAGA CCAATGATAA TTTAATAATG CCACCAACTC TGGCTTAGTT 660
 AAGTGAGAGT GTGAAGTGIG TGGCAAGAGA GCCTCACACC TCACTAGGTC CAGAGAGCCC 720
 AGGCCTTATG TTAATAATCAT GCACTTGAAA AGCAAACCTT AATCTGCAAA GACAGCAGCA 780
 25 AGCATTATAC GGTCACTCTG AATGATCCCT TTGAAATTTT TTTTGTGTTT GTTTGTTTAA 840
 ATCAAGCCTG AGGCTGGTGA ACASTAGCTA CACACCCATA TTGTGTGTTT TGTGAATGCT 900
 30 AGCTCTCTTG AATTTGGATA TTGGTTATTT TTTATAGAGT GTAAACCAAG TTTTATATTC 960
 TGCAATGCCA ACAGGTACCT ATCTGTTTCT AAATAAACT GTTTACATTC ATTATGGGGT 1020
 ATGTATGACC TTCATTTTCC AAGAAATAGA ACTCTAGCTT AGAATTATG ATGCTCTAAA 1080
 35 ATGTGAGAAAT GCGAACTCTC CTCGAAGTTC TCCCAAAC TCAGACAGCA CTCCTTCTC 1140
 CTAAATGAAT ATTCTTTTCT CCGTGTMTTC TGGTATTTTC TAGGCATCCT TCTCACCACA 1200
 40 GCCATAACCC TTTTFACTT CATTAGGCC GTATAACTGG NCGGACNGCT GGTGGGTATA 1260
 TAATACTGGT WCCAACAMAG GGTCTCTGGA TGTACACMAG GTTATCTT 1308

45

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 1705 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

TGGGCATGGA AGCGCTAGAA GGTTAGATT TTGAAACAGC AAAGAAGGAT TTCCTTGGAT - 60 -

CTGAGACCCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA 120

60

	TCAAGGAGCC CAAGGCGGCC GTGAGATGT ACATCTCAGC AGGAGAGCAC GTCAAGGCCA	180
	TGGAGATCTG TCGTGACCAT GGCTGGGTG ACATGTTGAT CGACATCGCC CGCAAACTGG	240
5	ATAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGGTACCTA CCTCAAGAA3 CTGAGACGCC	300
	CTGGCTAIGC TCGTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCTCTG GTGAGCTGC	360
	ACTGAGAGCC CAGCGCTGGG ATGAGGCTTT TGCTTTGCTT GAGAAGCATC CTGAGTTTAA	420
10	CGATGACATC TACATCCCGT ATGTTCACTG GCTAGCAGAG AACGATCCTT TTSAGGAAGC	480
	CTAGAAAGCG TTCTATAAGG CTGGCGGACA GAGAGAAAGG GTCTAGCTTC TTSAGCAGCT	540
15	CACAAACAAAT GCTGTGGCGG AGAGCAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT	600
	GTCCATGACG TGCTTGATA TAGCTCAAGA TCGTCCGAG AAGGACACAA TCGTTGGCAA	660
	TTTCTAGCAC TTGAGCGTTT TGGCAGAGCT GTACCATGCT TACCATGCCA TCCATGCCA	720
20	CACGGAAGAT CCGTTCACTG TCCATCTCC TGAAGCTTT TTCAAGATCT CCAGGTTCTT	780
	CTCTGACAC CTGCGCAAGG ATACGCGCTC GGGCATCTCT AAAGTGAAAA TACTCTTCAC	840
25	CTTGGCCAG CAGAGCAAGG CCGTCGCTC CTACAGGCTG GCGCGGAGG CTTATGACAA	900
	GCTGCTGCGC CTGTACATCC CTGTCAGATT CCAAAAGTCC ATTGAGCTGG GTACCTTGAC	960
	CATCGCGCGC AAGCGCTTCC AGGACAGTGA GGAGTIGTG CCGTTGCTCT ACCGCTGCTC	1020
30	CACCAACCAAC CGGCTGCTCA ACAACCTGGG CAACGTCTGC ATCAACTGCC GCGAGCCCTT	1080
	CATCTTCTCC GCGCTTCTCT ACGAGTCTCT ACACCTGCTT GAGTTCTACC TGGAGGAAGG	1140
35	GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGAG GTGCTGAGAC CCAAGCGGGA	1200
	TGACAGACAG CTAGAGATTT GCAAAACACA GCTCCGAGAT TCTTGGGGCT AGTGGAAGAC	1260
	CAAGCGATC CATCGAGAT NAGGACCGT TCACAGCTAA GCTTAGCTTT GAGCAAGGTG	1320
40	GCTCARAGTT CGTGGCACTG GTGCTGAGCC GCGTGGTCT GCGCTGCATG AGCGCGCGGG	1380
	ATCTCTCAT CAGGAGATGG GCGGACCCC TGAGGTGGCA ATACTTCGG TCACTGCTGC	1440
45	CTGAGCTTC CATTACCATG TCTCTCTCT GCTTCAGAT GTTCATCTT GAGGACTATG	1500
	ACTTCTGCT GTTTCAGAT GCTGCTGGC CTTACTGCG CAGGTGCAAG CATGACCTG	1560
	GCGCATGACC AGCATCTCT GAGGGGCTG CACCTCTGC GCGCTTGGG GTCTGCTGGG	1620
50	CTGTGAAGGA GAATAAGAG TTAACTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1680
	AAAAAAAAA AAAAAAAAAA AAAAA	1705

470

(A) LENGTH: 999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

	AGCAAATCAC CTTAAGGATC TGGAAATGAAA CTGTGACCAG TGTCGCCCTG GGTGGTTCTG	60
10	GAGAGACTGC CGTCTTCTTG TTGGGCGATA GGTGCTGGG CCCGGGCTTC AGTCACTGTC	120
	TCAGACAGKA GTCCCGATAA GCAGATCACC AGTCTCCAC TGTCTTCCT GTCGGCCTTG	180
	CTGCATGAGA AGATAGCTGC TTCCTCCCTC TTTTCTTACA CTGTAAATTA TTGTTTTACA	240
15	ATTGAGTGYC TTAATAATAG TYTACAAATA CTATGTATTT ATGCAAAACT GTTAAAGTTC	300
	TCATCTGTTA TGATTGGATA CTTGGTCTTG TCAGTAGTGG TCAGCATTCG GTTGTGAGCT	360
20	TGTCTACTC CATAGCTGT TATCCTGCTA TGCATTTTAC ATTGTGTGTT CACATCTATT	420
	CCAAAGAGGC TTCTAGAAA CAACATGGC GGTTCCTGCA GGCAGGCAG GCATTGCCCC	480
	ATGCTGTGT CCATAGGAGC CAATGAAAAG AACGTAGCTT GGTCTCTAG CCAGCCGTGG	540
25	GGTGGCGCAG GGCAGGCAGC CTCTGCACCA GACTCCAGTA CCGCCCATTT CCCCAGTCAC	600
	ACAATCATAC TCTTCTTTCA TAGAGATTTT ATTACCACCT AGACCACCCT AGTTTTCCTC	660
30	TCTGTAGTG TCTTAGCTC TTTTGCAACA AAATGTAGST ACAGACCAAT CCGTGTCCCT	720
	TCCCCAATCA GGAGTTCAC ACCATGAGTT GTTGGTTTTT CCAGAAGCTG CCAGTGGGTT	780
	CCCGTGAATT GCGTAAGAT ATCGATGATK TTTTATTG TTTTCTTCT TGTTTTTTTA	840
35	AATAATATAT TTAAGGCAG TATCTTTTGT ACTGTGAATT TCAGTAGAA GATGCAAAAT	900
	GCACTTTTTT TTAATTCTG TTGGTGTGTA TTGTATATAG TGTGTGTGCT TCTGTGATG	960
40	AAAATAAACT TTTTCTTTAT AAAAAAAAAA AAAAAAAAAA	999

45 (2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

55	GGCAGGAGTA GCATTTTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCATGT	60
	GATGTCTCA GCGAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA	120
60	GGCTGGAGAG ATCATATTTT TGGTATPAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA	180

471

TGTCTCTGT AGCAAACCGG AAAAGTCAAT GACAGAAGAT GCTGCTAGCG GTTTGAGCCA 240
 GAGAATGACA GCTGTGTTTT GGAGAAAAGG GCGGATGCT GCTCTAGAA AGCCCATCTT 300
 5 TCTGCTCTTC TTTTCTCTCC CCTTATATT GTGCTTTCAT TCTTTCATTC ATTCATCAAA 360
 CATTCTTTGA GCACCTATTA TGTGTCAAGC TCTGTCTAG CCTCTGAAA ACCTGCCCTC 420
 ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGAATACACT ATTATAAGCA CGGTTGTCA 480
 10 GGGTCTCACA GAGCAGTGGT CCTCATCCA GACCGATGAG CTAAAGAAG GCATCCAGGC 540
 GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC 600
 15 TGAAGTGGC AGTCTCTGA GTCTTGATTC CAGTAGAGG AGAGCAGTCT GTGAAAAGG 660
 ACCAAGGGTG GGAGAGGGA GAGCACATGG AGGAACTTCA GGTAGTTCTG GATGGCCTG 720
 GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATCTTCC TTGAGTTACA TGAAGTTCCA 780
 20 TCCCAATAAA CCCATTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTTA GGTGTGTCG 840
 AGTAGAATGA TTTTACAAAC GAATGATCA CAACAGTGA GAGATGTCTT TGTCTCTCT 900
 25 CCACTCCAC TGCTTCACT GACTAGCCTT TAAAAAAA A 941

30 (2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

40 TAAGTGAAT CCCCCGGT TGCAGGAAT TCGGCACGAG GCATTCTGAG AAGCTTAAGA 60
 CATACTTTGA AGACAAGCT AGGACCTCC AGCTCTGCG GCATGACCTA CTTTGCACC 120
 CCGAGTGGT GAAGCCCCAC CTGGGCATG TTCTGACTA CTTGTTCTT CTTCTCTCC 180
 45 GTGGCCTGGT TCGCCCTCAU AAGAGCGGA AGAAGCTGT TTCTCTTGT AGAAGGGA 240
 AGAGAGCAA GTCACAGAC CCACTGCGCA GTTCAAGCA AAGAGGAAAG AATTCAGAC 300
 50 CCACAGCCAA GCTCTCTGA GGTGTTGGG CTTCTCTGA GCTGAGCACA TTGTGAGCA 360
 CAGGCTTACA CCTTCTGTG AGAGGCGAG CTCTGTGCT TACTGCACAG CTTGAACAG 420
 CAGGCTTACA CCTTCTGTG AGAGGCGAG CTCTGTGCT TACTGCACAG CTTGAACAG 480

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3018 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

5	GCCAGCCTTA CAGSTTTTAC GTGAAATGAA AGCCATTGGA ATAGAAGCCT CGCTTGCAAC	60
15	ATATCACCAT ATTATTGGCC TGTTCGATCA ACCTGGAGAC CCTTTAAAGA GATCATCCTT	120
	CATCATTTAT GATATAATGA ATGAATTAAT GGGAAAGAGA TTTTCTCCAA AGGACCCGGA	180
20	TGATGATAAG TTTTTCAGT CAGGCATGAG CATATGCTCA TCTCTCAGAG ATCTAGAACT	240
	TGCCTACCAA GTACATGBCG TTTTAAAAAC GGGAGACAAC TGGAAATPCA TTGGACCTGA	300
	TCAACATCGT AATTTCTATT ATTCCAAGTT CTTGATTTG ATTCTCTAA TGGAAACAAAT	360
25	TGATGTACCG TTGAAGTGGT ATGAGGACCT GATACTTCA GCTACTTTG CCGACTCCCA	420
	AACAATGATA CATTTCTCTG AAGCATGGA TGTGGCAAT GGGCTAGAAG TGATTCTTAA	480
30	AATTTGGGAA AGATACTAAA GAATATGCTC ATACTTTCCG CAGTGACCTG AGAGAAGAGA	540
	TCTGATGCT CATGGCAAGG GACAAGCAAC CACCAGAGCT TCAAGTGGCA TTTGCTGACT	600
	GTGCTGCTGA TATCAAATCT GCGTATGAAA GCGAAGCCAT CAGACAGACT GCTCAGGATT	660
35	GGGCAGCCAC CTCTCTCAAC TGTATAGCTA TCTCTTTTTT AAGGGCTGGG AGAACTCAGG	720
	AAGCCTGGAA AATGTTGGGG CTMTTCAGGA AGCATAATAA GATTCTTAGA AGTGASTTGC	780
40	TGAATGAGCT TATGACAGT GCAAAAGTGT CTAACAGCCC TTCCAGGCC ATTGAAGTAG	840
	TAGAGCTGGC AAGTGCCCTC AGCTTACCTA TTTGTGAGGG CCTCAGCCAG AGAGTAATGA	900
	GTGATTTTGC AATCAACCAG GAACAAAAGG AAGCCTAAG TAATCTAACT GCATTGACCA	960
45	GTGACAGTGA TACTGACAGC AGCAGTGACA GCGACAGTGA CACCAGTGAA GGCAAATGAA	1020
	AGTGGAGATT CAGGAGCAGC AATGGTCTCA CCATAGCTGC TGGAAATCACA CCGAGAGACT	1080
50	GAGATATAAC AATATTTAAC ATTGTTACAA AGAAGAAAAG ATACAGATTT GGTGAATTTG	1140
	TTACTGTGAG GTACAGTCAG TACACAGCTG ACTTATGTAG ATTTAAGCTG CTAATATGCT	1200
	ACTTAACCAT CTATTAATGC ACCATTAAAG GCTTAGCATT TAAGTAGCAA CATTGCGGTT	1260
55	TTGAGACACA TGSTGAGTGC CATGGCTCTT GTCATAGGA TAAGCCTGCA CACCTAGAGT	1320
	GTGGTGAAGC TGACCTCAGG ATGCTGTCTT CGTGCSATTG CCGTCTCTCTG CTGCTGGACT	1380
60	TCTGCTTTG TTTGGCTGAT GTCTGCTGT GATGCTGCTC CTTCATCTTA GGTGTTGATG	1440

	CAGTTCTAAC ACAGTTGGGG TTGGGTCAAT AGTTTCCCAA TTTCAGGATA TTTCGATGTC	1500
	AGAAATAACG CATCTTAGGA ATGACTAAAC AAGATAATGG CAGTTTAGGC TGCACAACTG	1560
5	GTAAATGAC TGTAGATAAA TGTGTAAAT AGTGACAG TTGTATTTT TGTAAATATA	1620
	GGCGCTGCGA TAGTTTCTA ACTTGAACAG CATTGAATGT TTTATTTCTT CTTTPTTTTT	1680
	TTGTCTATAG CTGTTACCTA TTTTAGTGGT TGAAATGAGA GCTAGTGATG ACAGAAGGAT	1740
10	GTGGAATGTC TTCTTGACAT CATTGTGTAT TGCTGTAAAT CAAGTTGATA ACCACTACTT	1800
	CTAGCAGCTC TTACCCTAT GACTTAAGTG GTCTTGAAG GCAGTAAGTG GAGTTTGTGA	1860
15	GCATCTCTGC CTTCATGAGG GCTTCTACCA CTGACCTACTT TGCAGTAACC TGGCTCCCG	1920
	ATTTACTTAG GTACCCGAG AGTGTCCAC ATAAGCAGCT TGTCTTTAC CTGTCAGAG	1980
	TTGACAATTA TGGGATACTE TAGTCTACTT ATACTTGTGT TCCCATCTGT CTGCCATCTT	2040
20	CTCAAGCTTA GGACCCAGT ATACATCCTT AGAAACCAA GATCGTTT TGTPTTCTCT	2100
	TGGAATCTCA GGTCTTAAGG CATTTAATTG AGGGACAAA AAAAAAAAAA CCGGATATAG	2160
25	TAGCTAGCTA CTTAAGCATC CATGGGTATT GCTCCATATC AAACGAGATT TGCAGGACAG	2220
	AAAGAGTAAA TTAGCCTTCA GTCTTGGTTT ACAGCTTCCA AGGAGAGCCT TGGCACCTG	2280
	AAATGTTAAC TCGGTCCCTT CCTGTCTCTA GTTCATCAGC ACCTGCAGAT GCTTGACTCT	2340
30	TGTTAGCCTT ACTATTCAT ACASTCCTTA GATTACGGT ATGCTCTCTC CTATCCAGGC	2400
	ACCTATTCTG AATCACCATG TTGCTCTGCA GCTAAGTTG ATAGGAGAAA ATTCATTTG	2460
35	GTAGATGCC TATGAATTTG TAGTAGACTT TCAAAATGAG TGATTTGTTA GCTTGGTACT	2520
	TTTAAGTTTG TGSTACAGAT COTQCAAACC CATACTCTGA GCAATTAAT GCTTTGAACA	2580
	TAGAGAAAAA TTAAGGCCCTC ACAGGATGAG TGTCCATTCT CTGTAAATGC TTATTTTATC	2640
40	ATAGTCTTTA GCTCTAACT ATGAGTAAAA TTTTCTCTC GCGCGGGTGT GGTGACTCAC	2700
	ACCTGTAAAC TCAGCACTTT GCGAGGCAGA GTTGGGAGGA TCATTAGGT CAGGAGTTG	2760
45	CAGACTAGCC TGGCAACAT AGTGAGACAC CGGATCTAGA AAAAAATAAA AAGGCAGACT	2820
	GCTGTATGT ATCTGTCTC CAGTAATG GAGCGTGAG ATGGAGGAT TTTTGAAGCC	2880
	TAGGAGAAGG AGGTTGAGT GAGCCGTGAT CGCACTACTG CACTCCAGCC TGGCAACAG	2940
50	AGCAAGAGCC TGTCTTGGAG AAACGAGAAT TTTGGAAGAG CAAATGGGCG TGASTGCAST	3000
	GCCTCATGCC TGTAAATCC	3018

474

(A) LENGTH: 968 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

	GGCACGAGGG CCGCGGACACA TCCACGGGBC GCGAGTGACA CCGCGGAGGG AGAGCAGTBT	60
10	TCTGCTGGAG CCGATGCCAA AAACCATBCA TTTCTTATTC AGATTCATTG TTTCTTTTA	120
	TCTGTGGGBC CTMTTACTTG CTGAGAGACA AAAGAAAGAG GAGAGCACCG AAGAAGTGAA	180
15	AATAGAAATT TGCATGCTC CAGAAAATG CTCTAAGACA AGCAAGAAAG GAGAGTACT	240
	NAAATGCCA TTATGACGGC TACCTGGCTA AAGACGGCTC GAAATTCTAC TGCAGCGGA	300
	CACAAAATGA AGGCCACCC AAATGGTTTG TTCTTGCTGT TGGGCAAGTC ATAAAAGGBC	360
20	TAGACATTGC TATGACAGAT ATGTGCGCTG GAGAAAAGGG AAAAGTAGTT ATACCCCTT	420
	CATTTCATA CGGAAAGSAA GGTATGAG AGGCAAGAT TCCACCGAT GCTACATTGA	480
25	TTTTTGAGAT TGAATTTTAT GTGTGACCA AAGGACGAG GAGCATTGAG ACATTTAAAC	540
	AAATAGACAT GCACAATGAC AGGAGCTCT CTAAAGCGGA GATAAACCTC TACTTGCAAA	600
	GGGAATTTGA AAAAGATGAG AAGCCAGCTG ACAAGTCATA TCAGGATGCA GTTTTAGAAG	660
30	ATATTTTTAA GAAGAATGAC CATGATGGTG ATGGCTTCAT TTCTCCCAAG GAATACAATG	720
	TATACCAACA CSATGAACTA TAGCATATTT GTATTTCTAC TTTTTTTTTT TACCTATTTA	780
35	CTGTACTTTA TGTATWAAAC AAATCMCTT TTCTCCMAST TGTATTGCT ATTTTTCCGC	840
	TATGAGAAGA TATTTTGATC TCCCAATAC ATTGATTTTG GTATAATAAA TGTGAGGCTG	900
	TTTTGCAAAC TTAAAAAAA ATTTAAAAAA ACTGAGGGG GGCCCGTACC CAANTCCCG	960
40	NATATGAT	968

45

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1404 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

55	CGTTTTCCGG CCGTGCGTTT GTGGCCGTCC GGCTCCCTG ACATGCAGCC CTCTGGACCC	60
	CGAGGTGGA CCCTACTGTG ACACACCTAC CATGCGGACA CTCTTCAACC TCCTCTGGCT	120
60	TGCCCTGGCC TGCAGCCCTG TTCACACTAC CCTGTCAAAG TCAGATGCCA AAAAAGCCGC	180

CTCAAAGACG CTCTGGAGA AGAGTCAGTT TTCAGATAAG CCGGTGCAAG ACCGGGGTTT 240
 GGTGGTGACG GACCTCAAAG CTGAGAGTGT GGTTCCTTGA CATCGCAGCT ACTGCTCGGC 300
 5 AAAGGCCCGG GACAGACACT TTCTGGGGA TGTACTGGGC TATGTUACTC CATGGAACAG 360
 CCATGCTAC GATGTCACCA AGSTCTTTGG GAGCAAGTTC ACACAGATCT CAGCGGTCTG 420
 GCTGCAGCTG AAGAGACGTG GCGGTGAGAT GTTTGAGTTC ACGGGCTCTC ACGAGGTGGA 480
 10 CCAAGGGTGG ATCGGAGCTG TCAGGAAGCA TGCCAAGGGC CTGCAATAG TGCTCGGCT 540
 CCTGTTTGA GACTGGACTT ACSATGATTT CCGGAACCTC TTAGAAGTG AGGATGAGAT 600
 15 AGAGGAGCTG ACCAAGACCG TGTTCAGGT GGCAAAGAAC CAGCATTTCT ATGGCTTCTT 660
 GGTGGAGCTC TGGAACTAGC TGCTAAGCCA GAAGCGGCTG GCGTATCTC ACATGCTCAC 720
 CCACTTGCCG GAGGCTCTGC ACCAGGCCCG GCTGCTGGCC CTCTTGTCA TCCGGCTCC 780
 20 CATCACCCCT CCGACGACG AGGTGGGCTT GTTCAGGAC AAGGATTTG ACAGGTGCG 840
 CCGCGTCTG GATGTTTCA GCTCATGAC CTACGACTAC TCTACAGGGC ATCAGCCTGG 900
 25 CCTAATGCA CCGCTGTCTT GGTTCGAGC CTGCGTCCAG GTCTTGAGC CGAACTCCAA 960
 GTGCGGAAGC AAAATCTCTC TGGGCTCAA CTCTATGCT ATGACTAGG CGACTCCAA 1020
 GGATGCCCCG GAGCTCTTTG TCGGGGCCAG GTACATCCAG ACACTGAAG ACCACAGGCC 1080
 30 CCGGATGGTG TGGACAGCC AGGYCTCAGA CCACTTCTTC GAGTACAAGA AGAGCCGCA 1140
 TGGAGGGCAC GTGCTCTTCT ATCCAACCTT GAAGTCCCTG CAGGTGCGGC TGGAGCTGGT 1200
 35 CCGGAGCTG GCGTGTGGG TCTCTATCTG GAGCTGGGC AGGGCTTGA CTACTTCTAC 1260
 GACCTGCTCT AGGTGGGCT TCGGCTCTC GCGGTGAGC TGTCTTTTC TAAGCCATGG 1320
 40 AGTGAGTGAG CAGGTGTGAA ATACAGCCCT NCACTCCGTT TGCTGTGAAA AAAAAAAAAA 1380
 AAAAAAAAAA AAAAAAAAAA AAAA 1404

45

(2) INFORMATION FOR SEQ. ID NO: 213:

50

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 707 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ. ID NO: 213

60

CATGAGAGCTG ATCAAGGAGG AGCTTCCCAT GAGCATCAAG GATCAAGAG GAAATTCAG 1460

5 CCGCTGCATC GCAGACGTGG TCTTCTCTTT GATGCGGTG ATGACAGAGC TCGGCTTGA 240
 GATCGCGCGC ATGATGAGAG TCGAGCGGA CTTGCGAGAG CTGATCGAGA CCAATGACCG 300
 CATGAGCGAC CTGCGACCGG ACTTCTAGGG CCGCGAGAG GTGAGTCAAT GCGTGCAGAC 360
 CCTGAGCGGG ATGTGGGGGT CAGATGAGTT GAGCGACTCA CAGGTGCTTC AGATGCTGTT 420
 10 CGAAGCTGAG TAGGCTTACA ACGCTTCAA CCGCTGCTG CAGGCTGAG CCGCGCGGAC 480
 TAGGCTTGG ACAGAAGGG AGATCTGAG CGCATGGGTG CTGCTGCTCT GTGCGGACA 540
 CAGGCGGTGG TCATGCACAC AATGCACTGT TTGAGCTGG TGTCTGCTG TGTGCTGTTG 600
 15 GTGTGAGAAC TTTTGGGGG GGGCGCTGG CACAATGAG ATGCTCTTGG ACCTTCAAAA 660
 AAAAAAAAAA AAAAAGTGG GGGGGCGGG GTGCAATGG TCGCTTCT 720
 20

(2) INFORMATION FOR SEQ ID NO: 224:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1384 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

GGGGAATTGC AGTGACAGCA GGATTAAGAG TCGAGCGGAG GACAGAGCTG GGACACAGGT 60
 35 ATGAGAGAGG GGTTCAGGGA GGTAAAGAG GCGAGACTAT CAGGCTGCGG GCGGTGAGAA 120
 TCGAGGAGAG GGAGCGGAAA CAGAAAGAGG GAGAGAGACC GGGGCACTTG TCGGTTGAG 180
 40 AGCGGCTGAG CCATGTTGGG AGGAAAGGGA TACTGGCTAG CAGGCTGCTT ACAGAGTGG 240
 GGGCTGCTCT TGGTTCTGCT GCTTCTGCTG CTGGGCGGGG GGTGGGCGCA GGAGGGGTCA 300
 GAGCGGCTGC TGCTGAGAGG GAGTGGCTG GTGCTCTGTA AGGCTGCGGG AGCTGCTGCA 360
 45 GGGGGGCGGG GGGGAGCAG CTTGGGAGAG GCGCGGCTG GGGGAGTGGC ATTTGCTGG 420
 GTGGAAGGCG AAGACCATGA GCGAGCAGGG GAGACCGGGA ATGGGACCAK TGGGCGCATC 480
 TACTTGAGCC AGGTCTGTGT GAACGAGGGC GGTGGCTTTC AGGGGCTTTC TGGCTGCTTC 540
 50 GTAGCGGCTG TCGGGGTGT CTACAGCTTC GGGTTCATG TGCTGAGGT GTACAACCGC 600
 CAAACTGTCT AGGTGAGGCT GATGCTGAAC AGGTGGGCTG TCATCTCAGC CTTTGCGAAT 660
 55 GATCTGAGG TGACCCGGGA GGGAGCCAG AGCTCTGTGC TACTGCGCTT GGAGCTGGG 720
 GACCGAGTGT CTCTGCGGCT GCGTGGGGG AATCTACTTG GTGGTTGGA AATCTCAAGT 780
 60 TTCTCTGGCT TCTCATCTT CCGTCTCTGA GGACCCAAAT YTTTCTAGCA CAGAAATCCA 840

GGGCCCTBACA ACTTTCCTTCT GCGCTCTCTT GCGCCAGAAA CAGCAGAGGC AGGAGAGACA 900
 CTGCTCTGG YTCCTATCCC ACYTCTTTGC ATGGGAMCCT GTGCCAAACA CCCAAGTTTA 960
 5 AGARAAPARY ARARCTGWSG CAGGTATACA GAGCTGGAAG TGGACCATGG AAAACATSCA 1020
 TAACCATGCA TCTTCTTGGT TGGGACCTC CTGAAACTGT CCACCTTTGA AGTTTGAAC 1080
 TIAGTCCCTC CAMACTCTGA CTGCTGCTC CTTCCTCCCA GCTCTCTCAC TGAGTTATTT 1140
 10 TCACTGTACC TGTTCCAGCA TATCCCACT ATCTCTCTTT CTCTGATCT GTGCTGTCTT 1200
 ATTCTCTCC TIAGGTTCG TATTACCTGG GATTCATGA TTCATTCTT CAGACCTCT 1260
 CCTGCCAGTA TGCTAAACCC TCGTCTCTC TTTCTTATCC CGCTGTCCCA TTGGCCCAAC 1320
 15 CTGGATGAAT CTATCAATAA AACAACTAGA GAATGGTGGT CAAAAA AAAA AAAAAC 1380
 TCGA 1384
 20

25 (2) INFORMATION FOR SEQ ID NO: 225:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 750 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

GGGTCGACCC ACGGCTGGC TGACCACTC GTTATAGATA CTTCTTCTA TACCAAACT 60
 35 GTTTAAACAG GTGACACAC AAGGGATGT GTCTTACTC TCTGCGGTC TTCAAGCATC 120
 CCTTTGTGGG AAATCTCTT GGCACAGCA GTGCTATTG GTCTGCTCT TCTTCTCTT 180
 40 TTTCACCCAG GGATGTTGT ATCATAAGTC AAAACAACAG TATATTCAA ATCTCAAAAG 240
 CTATTGTGAC CTGAGCACA TTGAAATCT GCAGAGTTT TCTATGTAG CTTTAGASTA 300
 ACTCTCTGC TTCTCTGTA CTTACAATC AGGTTCTGC TTTCCTAAG AGCATGAGCA 360
 45 GAAGATCTT CATGTGACG CTAGTTCTAT TGCAGTCTG GGTGAAACTA TTAAAGCAT 420
 GGGGTGCTK CTCCCAAT CTCCCTAAC AATTCGTTT GTGGACTTCT CATCTAAAAG 480
 50 GTTAGTGGCT TTGCTTGGG ATCAGTCTC TCTATTGATG TTCTTGCTG TCTCCAGACA 540
 CATTCCTGTT GCATTAGAC TTGAAAGACT TGTAGATGT TGATGTTTCA GCACAGGATG 600
 TCTCTCTCTT TCTCTCTCTT TCTCTCTCTT TCTCTCTCTT TCTCTCTCTT TCTCTCTCTT 660

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2057 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

5	CCGAGCCCGGC TGCGCCGGGG GAATCCGTGC GGCGCCCTTC CGTCCCRGTC CCATCCTCGC	60
15	CGCGCTCCAG CACCTCTGAA GTTTTGCAGC GCGCAGAAAG GAGGCGAGGA AGGAGGGAGT	120
	GTGTGAGAGG AGGGAGCAAA AAGCTCAGCC TAAAACATTT ATTTCAAGGA GAAAAGAAAA	180
20	AGGCGGCGCG CAAAAATGGG TGGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC	240
	ATTGTGTGTC GGATTCTGCT CTTGTTCGAA ATCTCGGCTT TTCTGGTGCG AGGCTTGATT	300
	GCTCCAGGGC CCACAACGGC AGTGTCTTAC ATGTGGGTGA AATGTGTGGA TGCCCGTAAG	360
25	AAGCATACAA AGACAAAATG GTTGTGTGCT TGCGGACCCA ATCATTGTGA CAAGATCCGA	420
	GACATDMAAG AGCAATTCC AAGGGAAATT GAAGCCAATG ACATCGTGTT TTCTGTTCAC	480
30	ATTCCCTCCG CCCACATGGA GATGAGTCCT TGGTTCCAAT TCATGTTGTT TATCCTGCAG	540
	CTGGACATTC CTTTCAAGCT AAAAACCAGC ATCAGRGAAA ATGCAGAAAT CTCCATGGAC	600
	GTTCCTCTGC CTTACCGTGA TGATCCGTTT GCTGAGTGA CTGAAATGGC CCATGAAAGA	660
35	GTACCAAGGA AACTCAAATG CACTTTCACA TCTCCCAAGA CTCCAGAGCA TGGAGGGCGG	720
	GTTACTATGA ATGTGATGTC CTTCTTTTCA TGGAAATPAG GTCTGTGACC CATGAAGTTT	780
40	TACCTTTTAA ACATCCGGCT GCGTGTGAAT GAGAAGAAGA AAATCAATGT GCGAATTGGG	840
	GAGATAAAGG ATATCCGGTT GGTGGGATC CACCAAAATG GAGGCTTCAC CAAGGTGTGG	900
	TTTGCCATGA AGACCTTCCT TACGCCAGC ATCTTCATCA TTATGGTGTG GTATTGGAGG	960
45	AGGATCACCA TGATGTCCCG ACCGCCAGTG CTTCTGGAAA AAGTCATCTT TGCCCTTGGG	1020
	ATTTCATGTA CTTTATCAA TATGCCAGTG GAATGGTTTT CCATCGGGTT TGA CTGGAAC	1080
50	TGGATGCTGC TGTTTGGTGA CATCCGACAG GCATCTTCTA TGCATGCTT CTCTCCTTCT	1140
	GGATCATCTT CTGTGGCGAG CACATGATGG ATCAGCACGA GCGGAACCAC ATCGCAGGGT	1200
	ATTGGAAGCA AGTCGGACCC ATTGCCGTTG GTCCTTCTGC CTCTTCATAT TTGACATGTG	1260
55	TGAGAGAGGG GTACAACCTA CGAATCCCTT CTACAGTATC TGGACTACAG ACATTGGGAA	1320
	CAGAGCTGGC CATGGCTTTC ATCATCGTGG CTGGAATCTG CCTCTGCCTC TAACTTCCTG	1380
60	TTTCTATGCT TCATGGTATT TCAGGTGTTT CGGAACATCA GTGGGAAGCA GTCCAGCCTG	1440

CCAGCTATGA GGAAGTGGG GGGCTAGAG TATGAGGGG TAATTTTTAG GTTCAAGTTG 1500
 CTCATGCTTA GCACTTGGG CTGGCTGGG ATCACTCTTA TCTTCTTCAT CGTTAGTCAG 1560
 5 GTAACGGAAG GGCATTGGGA AATGGGGGG CGTCACACT GCAAGTGAAC AGTGCCTTTT 1620
 TCAAGGCAAT CTATGGGAG TGGATCTGT AGCTCTTTC TGTGATGTTG TTGTATGCAC 1680
 CATGCCATAA AACTATGGA GAGACCACT GCAAGCAAT GCAACTCCCA TGTAAATCGA 1740
 10 GGAAGAGTG TCTTTTGTT GTTGGGAGC TTATGAGA ACTGTTGAG GCTTCGAAAT 1800
 ATTCTTCAT CAATGAGAG GAGCTCTTG GATCTCAAT CAACAAGGCA ACACATGTTT 1860
 15 ATCAGCTTTC GATTGGCACT TTTGAGATC AGCTGATTC TACTTGTATA GGCACACAAA 1920
 TACTCTCAT TACCTCTTA CTCAAACTT TAAATATAA GAAAAAGCG TCAACAATAA 1980
 ATATCTTTC AATATCTCT TACTCTCTT AAAAAAAA AAAAAACTC GTGACGAATT 2040
 20 GGGAGAGAG GAGACA 2057

25

(2) INFORMATION FOR SEQ ID NO: 227:

30

(1) SEQUENCE CHARACTERISTICS:

- (a) LENGTH: 2064 base pairs
- (b) TYPE: nucleic acid
- (c) STRANDEDNESS: double
- (d) TOPOLOGY: linear

35

(2) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

40

GCGAGAGGG GATTCTTTC AAGAGGCCA ACCCTGATC CTGTGTGGC CTCTCTCCC 60
 ACCAAGTGT TTATAAAAA AGCTCTTCT ACCGAGATA ACTGTTGATT TTCTACTCT 120
 40 CCGTCTTAG TCACTTTT GAGAAAAA ATTGCATTC TGGAAACCAG AAGAAAAATA 180
 TCAAGCGGG AATGATCTG TATGTGCTT GCTGCTTTC GCTGAGTGT TGGAGTCTG 240
 CTCAGCTTT AATTAGCTG TCTTGATTC TCTGCTTTC AAGAGAACG CTGTTCAGA 300
 45 GCTGTGATG GGTTCGACT GAGAGAGAG TCTCTTGGG TCTTGTACT GCGGCGCTT 360
 CTCTCTCTT GATGATCCG AGCAGCACT CTGCGGAGG CAGAAGGTAC CGGGGCAGCT 420
 50 ACTCGAGGAG TGTGCGGGG TGGTGGGCT GCGGCTCTG CCGTGGGGG CTGTTGCTG 480
 TGTGATCTA TTCTACTAC TGGCTGAAA ATGCGTTCG CCGGCGCTT ACTTGGATG 540
 TCTGATCTA TCTGATCTA TCTGATCTA TCTGATCTA TCTGATCTA TCTGATCTA 600

60

AATTAGGAT AATTAGGAT AATTAGGAT AATTAGGAT AATTAGGAT AATTAGGAT

480

CTCTCCCAT TGGACTGTG GGTGCTGAT AACCTGAGTA TGGCTGACCC CAACATTGCG 840
 TTCTGTGATA AACTGCCCCA GCAGACCGGT GATCGTGCTG GATGUAAGGA TGGGTTTAC 900
 5 AGCAACAGCA TGTATGAGT TCTGAGAAC GGGCAGGGG GGGGCACTG TGTCTGAGT 960
 TACGCCACCC GCTTGCAGAC TTTGTTTGGC ATGTACAAAT ACAGTCAAGC TGGCTTTAGC 1020
 10 GGGGAGGATA GCTTGCAGCA GGCACAACTC TTCTGCGGA CACTGAGGA CATCTGGA 1080
 GATGCCCCCTG AGTCTCAGAA CAACTGCGGC CTCATTGGCT ACCAGGAAC TGCAGATGAC 1140
 AGCAGCTTCT GCTGTGACA GAGGTTCCTC GGGCACTGC GGCAGGAGGA AAAGGAAGAG 1200
 15 GTTACGTGCT GAGCTTTGAA GACCTGAGCG GTGCGGASTA GCTGCAAGAT GTTCAAGAG 1260
 GCTGAGCTCT GATCAGTGG AATGGAAGG GGGCTGCTG TGGGCAAGGA TTTCTCTGGA 1320
 20 GACCCAGGCT CAGCAGGGA GAGGCTCCAG TGCTGTGCAA GCTTGTGAG TGGGGCTCT 1380
 CTTCACTGCG TGAATGTGCA GCAGAGCTAT TTCTTTGCA AGGGGGCTTT GCAAGGAAGG 1440
 GTCCAGGACT TGACATCTTA AGATGCTCT TGTGCTGCTG GGCAGTAT TTGCTCTCTC 1500
 25 TGAGGCTGCG TGTCTTCAAC CTGTGAAATG GGATCATAAT CACTGCTTA GCTGCTCAC 1560
 GGTGTTGCTG AGGACTGACT GTGTGGAAGT TTTTCATAAA CTTTGGATG TATGTACTT 1620
 30 AGGGGCTGCT GAGGTGTCT TTCATGGGCG CTTCAGAGC CACTGCTCA GCTTCTGCCC 1680
 TTCCTTTGCT GGGGACGCG GAACTCTCTC AATGATATCA ACAGGCTGCT TGGGCTCTG 1740
 GCTCTGCTG ATGTGCAAT ATGCGGAGC CCCAGCAGAA GAATGGAGAG GAGGAGGAG 1800
 35 CTGAGTTTGG GTATTGAAT CCCCCGGCTC CCAGCTGCA GATCAAGGT TGATATGGAC 1860
 TCTCTGCGG GCAACTCTT GCTAATCAT GACTATCTCT AGGATTCTG CACCACTTCC 1920
 40 TTCCCTGGCC CTTAAGCCT AGCTGTGTAT CGGCACCCC ACCCACTAG AGTACTCCCT 1980
 CTCACTTGGG GTTCTCTTAT ACTCCACCCC TTTCTCAAGG GTCCTTTTAT AAAGCACATC 2040
 45 TCAGATTAAA AAAAAAAAAA AAAAAAAAAA AGGGGGGGCN GCNT 2084

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2143 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

TCGACCCACG CGTCCGGTTG AATTCCTTGA CCTGCAAACA CATATTTATT AGCCTGACTC 60

	AAACAATGAA GCTATTAAAA CTTCGGAGGA ACATTGTAAA ACTCTCTTTG TATCGGCATT	120
	TCACCAACAC GCTTATTTTG GCASTGGGAG CATCCATTGT GTTTATCACC TGACAAACCA	180
5	TGAASITCAG AATAGTGACA TGTCAGTCGG ACTGCGGGA GTTGTGGSTA GACCATGCCA	240
	TCTGGCGCTT GCTSTTCTCC ATGATCCTCT TTGTATKAT GCTTCTCTCG CCACCATCTG	300
	CAACAACCA GAGSTTTGCC TTTCACCAT TGTCTGAGGA AGAGGAGGAG GATGAACAAA	360
10	AGGAGCCTAT GCTGAAAGAA AGCTTTGAAG GAATGAAAAT GAGAAGTATC AAACAAGAAC	420
	CCAATGGAAA TATTAAGTT AACAAAGGAC AGGAGATCA TTGGAAGTGG GTAGAAGAGA	480
15	ATGTTCTTTC TTCTGTGACA GATGTAGCAC TTCCAGCCTT TCTGGATTCA GATGAGGAAC	540
	GAATGATCAC ATACTTTGAA AGGTCCAAAA TGGAGTAAGG AATGGGAAGA TTGCACTTA	600
	AAGATGGCTA CCAACAGGGA ACAGATCAGC ATCTTTGTCA GTCTTCTGTA CGGCTCCATG	660
20	GGATTAAAGG AACCAATGAC ATCCTGATCT GTTCTTTGAT CTTTGGGCAT TGGAGTTGGT	720
	GAGAGGTGTC AGAACAAGA GAACATCTTA CTGAAAACAA GTTCATAAGA TGAGAAAAAT	780
25	CTACGAGCTT CTATTTTACA ACACTGCTGC CCGCTTTCCT CCGAGACTCT GACATGATG	840
	TTCATGCCAA TTAAGTGTGT TGTTCCTGAA CTTTCTGTAA TGTTCATTT TTTAAATCTG	900
	ACAAACTAAA AAGTTTAACG TCTTCTAAAA GATTCTCATC AACACCATAA TATGTAATCT	960
30	CCAGGAGCAA CTGCCTGTAA TTTTATTTA TTAGGGAGT TACATAGGTG ATGGCGGAAA	1020
	TTGTAACTA CTTTTCATTT TCCTGGGAAG TCAAGGTTAC ATCTTGAGA GTTGTTTTTG	1080
35	AGAAAAAGG GCGCTTCTGA GTTAAGGAGC CATAGTTCTA TCAATGATCA AAAGAAAAA	1140
	AAAAAAAGA GAACTGTTA CAGTATGATT CAGATCATTT AAAAAAGCA AATCAAGTGC	1200
	AATTTTGTTC ACAAATGGTG TATATTAAAG ATTTTCTAT TTCAGATGTA CTTTAAAGAG	1260
40	AAATATTAGC TTAACCTTTT TGACATCTGC TATTGTGACA CATCCCATG CTGGCAATGT	1320
	GGTGACACT CCGAACTTT TAACTACTGT TTGTAAAGC TCCAAGGCTG GCATTGCAAG	1380
45	GTCTTAGGC AATGTTTTGT TTGCTTTAT GCAAGAGGT GCTCCAGTG TTGTCAATTA	1440
	GCACCGTGT AGAGGAAGTG TAATGCTTCA GAAATGTAG CTTATACAAA GGAAACAGGT	1500
	CCTGCTGGCT TAATTTAAAC AGTTATGCA TGAAGTAGG TGGAGGCTCT GCACTGCTGC	1560
50	TGTTTCTTA GGATGAGTG TTCTGGTATC TGGTATTGGT TTAGAGACTG TTAATAAGGG	1620
	TTTATTTTAA TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT	1680
	TTTATTTTAA TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT	1740
60	TTTATTTTAA TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT TTTTATTTT	1800

482

TGTAAAAACC AATGACACACA TACCACAAT CTTCACAAAC TCATACTACA GTGAAAGTGT 1920
 TAACCCCTTAG GTAGTTTCTC TACAACTCTT TGCTATGGTG ATTTTATAAA AAGTTTCTTA 1980
 5 GGGAAAGTATC TCTGAGGGAA CAGGCAATCT GAAGGAACTG ACTATATTCT CCATGGCTAA 2040
 GTCCATTAGG CCAAAAENCT GGGTGGGTAT TGTTTGTGAT GCTGTCTATT GGCATATTAA 2100
 10 AAACGTAGGC CGGANGBAAT AATTAGGTTG TNAATCCCGC GGG 2143

(2) INFORMATION FOR SEQ ID NO: 229:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1025 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

25

CCTGCCCCAG ATTTGCTTCAT TGAGCTGACC ATGCGCCTCT ACTATGCCAG CCGCTAGTCC 60

CTGACAACTT CCACCTTGAT TCGGACCTCT GTAGATTGAG CGCCACACCC AGATCCCGCT 120

CCCAGGCGTT CCGCTCTCTC CCATCAGCAG CCGTGTAAAC AGTGCTTCTT GAGAAAAGTT 180

30

GGAGAAGTGA GGGGAGCCAG GTTATTCTCT GGAGGTTGCT GGATGAAGGG GTAGCCTAGG 240

AGATGTGAAG TGTAGGTTTG GTTAAGSAAA TGTATTACAT CCCCCACCCC CAACCAAGTT 300

35

CTTCCAGACT AAAGAATTAA GGTAAACATCA ATACCTAGGC CTGAGAAATA ACGGATCTCT 360

TGTTGGGGAG CTGCTGCTT TGTCTGCTAT GAACAGAGTT BATGAAAGTG GGTGTGCGC 420

AACAAGTGGC TTTCTTGCC TACTTTAGTC AACCAGCAGA CCGACTGAG CTGCTAGTC 480

40

CAGCCCAGCC ATGTGTCATG ACTCTTCCAT AAGGGATCCT CACCCTTCCA CTTCATGCA 540

AGAAGGCCCA GTTCCACAGC ATTATACAAC CATACCCCA ACCACTCTGA CACTCTCTC 600

45

CAGTTCCAGC AATGCTTAGA GATATGCTCC CTGCTCTCTC CACAGTGTG CTCCACAC 660

CTAGCCTTTG TTCTGGAAAC CCGAGAGAGG GCTGCGCTT ACTCATCTCA GGSAAATGTAG 720

CCCTGCGGC CTGGCTTAAG CCGACACTCC TGACCTCTCT GTTCAGCTG AGGCTGTCT 780

50

TGAAGCCCGC TACCACTCT GAGGCTCCTA GGAGGTACCA TGCTTCCCAC TCTGGGGCCT 840

GCCCCCTGCT AGCAGTCTCC CAGCTCCCAA CAGCTGGGG AAGCTCTGCA CAGAGTGACC 900

55

TGAGACCAGG TACAGGAAAC CTGTAGCTCA ATCAGTGTCT CTTTAACTGC ATAAGCAATA 960

AGATCTTAAT AAAGTCTTCT AGGCTGTAGG GTGCTTCTTA CAACCACAGC CAAAAAAAAA 1020

AAAAA 1025

60

(2) INFORMATION FOR SEQ ID NO: 230:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1250 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

5 GGCACGGCT CCGCCACGC GTXGSGGT GCGAGTATG GGGCGGTGAT GGCATGGAG 60
 15 GGTACTGBC GGTTCCTGBC GTTCTTGGG TCGGCACTGC TCGTCGGCTT COTSTGGTG 120
 ATSTTCGCC TCTCTGSSG CTTCCAATAC CGAGAGGGGC TTGGCTGGGA TGGGAGGCA 180
 CTAGAGTTA ACTGGCAGCC ACTCTSATG GTTACCGGT TCGTCTTCAT CCAGGGCATC 240
 20 GCATCATCT CTACAGATG CCGTGGACCT GAAATGCAG CAAGCTCTG ATGAAATCCA 300
 TCCATGCAGG GTTAAATGCA GTTCTGCA TTCTTGAAT TATCTCTGTG GTGGCGGTG 360
 25 TTGAGAACCA CAATGTTAAC AATATAGCCA ATATGTACAG TCTGCACAGC TGGTTGGAC 420
 TGATAGCTGT CATATGCTAT TTGTTACAGC TTCTTTGAGG TTTTTCAGTC TTTCTGCTC 480
 CATGGGCTCC GCTTTCTCTC CGACCATTTT TCATGCGCAT ACATGTTTAT TCTGGAATTG 540
 30 TCATCTTTG AACAGTGATT GCAACAGCAC TTATGGGATT GACAGAGAAA CTGATTTTTT 600
 CCCTGAGACA TCTTCATAC AGTACATTCC CGCCAGAAGG TGTTTTCGTA AATACGCTTG 660
 35 GCTTCTGAT CTTGCTGTC GGGGCTCTCA TTTTTGCTAT AGTCACAGA CCGCAATGGA 720
 AACGTCTTAA GAGGCCAAT TCTACTATTC TTCATGCAAA TGGAGGCACT GAACAGGGAG 780
 CAAGAGGTTT CATGCGAGCC TACTCTGGCA ACAACATGGA CAAATCAGAT TCAGAGTTAA 840
 40 ACATGAAAT ABCAGCAAGG AAAAGAACT TAGCTCTGGA TGAGGCTGGG CAGAGATCTA 900
 CCATGTAAAA GTTCTAGAG ATAGAGCCAT ATACCTTAC GTTTCAAAAC TAGCTCTACA 960
 45 GTTTCTGCTT CCGATTAGC CATATGATAA TTGGGCTATG TAGTATCAAT ATTTAATTTA 1020
 ATCAGAAAG ATGTTTCTT GAAATAATTT GATTTSATG AGGCTATGA ACTACCTGA 1080
 ATTGAAAGG ATGTGATTAA TATAAATAAT AGCAGATATA AATGTGGTT ATGTTACCTT 1140
 50 TATCTTGTG AGGACCAGAA CATAGCAGG GTGCTTGTG CAAATAGAT ACTCAATATG 1200
 TGAATATGTC CTTATAGTA CTTAATGGA TAAATGGA GCATGCTGA 1250

(2) INFORMATION FOR SEQ ID NO: 231:

60 (i) SEQUENCE CHARACTERISTICS:

484

(A) LENGTH: 1811 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

	CNCGNCAGTAC CGGTGCGATT CCCGGGTGGA CCGACGGGTC CCGTGCATTG CAGGGCCTTT	60
10	CAGTGGGCTTT CATTCTGAAG TTCTTGGAAC ACATGTTCCA TGTCTGATG GCCCAGGTTA	120
	CCASTGTGAT TATCACAACA GTGTCTGTCC TGSTCTTTGA CTTCAAGCCC TCCCTGGAAT	180
15	TTTTCTTGA AGCCSCATEA GTCTCTCTCT CTATATTTAT TTATAATGCC AGCAAGCCTC	240
	AAGTTCCCGA ATACGCACCT AGGCAAGAAA GGAATCCGAGA TTTAAGTGGC AATCTTTGGG	300
	AGGTTCCAG TGGGGATGGA GAAGAACTAG AAAGACTTAC CAAACCCAAAG AATGATGAGT	360
20	CAGATGAAGA TACTTTCTAA CTGGTACCCA CATACTTTGC AGCTCTCTTG AACCTTATTT	420
	TCACATTTTC AGTGTTTGTA ATATTTATCT TTTCACCTTG ATAAACCAGA AATGTTTCTA	480
25	AATCTTAATA TTCTTTGCAAT ATATCTAGCT ACTCCCTAAA TGGTTCCATC CAAGGCTTAG	540
	AGTACCCAAA GGTAAGAAA TTCTAAAGAA CTGATACAGG AGTAACAATA TGAAGAATTC	600
	ATTAAATATCT CACTACTTGA TAAATCAGAA ACTTATATCT GCAGATTATT TTCCTTGGCC	660
30	TTCAAGCTTC CAAAAAAGTT GTAATAATCA TGTTAGCTAT AGTTTGTATA TACACATAGA	720
	GAICAATTTG CCAAATATTC ACAATCATGT AGTTCTAGTT TACATGCCAA AGTCTTCCCT	780
35	TTTTAACATT ATAAAAGCTA GGTGTCTCT TGAATTTTGA CGCCCTAGAG ATAGTCATTT	840
	TGCAAGTAAA GAGCAACGGG AUCCTTTCTA AAAACGTTGG TTGAAGGACC TAAATACCTG	900
	GCCATACCAT AGATTTGGGA TGATGTAGTC TGTGCTAAAT ATTTTGCTGA AGAAGCAGTT	960
40	TCTCAGACAC AACATCTCAG AATTTTAATT TTTAGAAATT CATGGGAAAT TGGATTTTGG	1020
	TAATAATCTT TTGATGTTTT AAACATTGGT TCCCTAGTCA CCATAGTTAC CACTTGATTT	1080
45	TTAAGTCATT TAAACAAGCC ACGGTGGGGC TTTTCTCTCC TCAGTTTGAG GAGAAAAATC	1140
	TTGATGTCAT TACTCTGAA TTATTACATT TTGGAGAATA AGAGGGCATT TTATTTTATT	1200
	AGTTACTAAT TCAAGTGTG ACTATTGTAT ATCTTTCCAA GAGTTGAAAT GCTGGCTTCA	1260
50	GAATCATACC AGATTSTCAG TGAAGCTGAT GCTAGGAAC TTTTAAAGGG ATCCTTTCAA	1320
	AAGGATCACT TAGCAAACAC ATGTTGACTT TTAAGTGATG TATGAATATT AATACTCTAA	1380
55	AAATAGAAAG ACCAGTAATA TATAAGTCAC TTACAGTGC TACTTCACAC TTAAAAGTGC	1440
	ATGGTATTTT TCATGGTATT TTGCATGCAG CCAGTTAACT CTCGTAGATA GAGAAATCAG	1500
	GTGATAGATG ATATTAAAAA TTAGCAAACA AAAGTGACTT CTCAGGGTC ATGCACTGG	1560
60	GTGATGATAG AAGAGTGGC TTTAACTGGC AGGCTGTAT GTTTACAGAC TACCATACTG	1620

TAAATATGAG CTTTATGGTG TCATTCTCAG AACTTATAC ATTTCTGCTC TCTTTCTCC 1680
 TAAGTTTCAT GCAGATGAAT ATAAAGTAAT ATACTATTAT ATAATTCATT TGTGATATCC 1740
 5 ACAATAATAT GACTGGCAAG AATTGCTGCA AATTGTATAT TAAATAAAT ATTAACCTA 1800
 AAAAAAAAAA N 1811

10

(2) INFORMATION FOR SEQ ID NO: 232:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2271 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CTGACCTCAT GCGTAGAGC CTAGTAACAG CCGAGGCTCC CAGCCGAGTC CGTTATGGCC 60
 25 GCTGCCCTCC CGAAGAGGAT GAGGGGGCCA GCACAAGCGA AACTGCTGCC CCGTCCGGCC 120
 ATCCAAGCCC TTGTGGGGTT GCGGGGGGCT CTGGTCTTGG CGCTCCTGCT TGTGTCCGCC 180
 GCTCTATCCA GTGTTGTATC ACGGACTGAT TCACCGAGCC CAACCGTACT CAACTCACAT 240
 30 ATTTCTACCC CAAATGTGAA TGCTTTAACA CATGAAAACC AAACCAAACC TTCTATTGCC 300
 CAAATCAGCA CCACCCTCCC TCCCACGAGC AGTACCAAGA AAAGTGGAGG AGCATCTCTG 360
 35 GTCCCTCATC CCTGGCCTAC TCCTCTGTCT CAAGAGGAAG CTGATAACAA TGAAGATCCT 420
 ACTATAGAGG AGGAGGATCT TCTGATGCTG AACAGTTCTC CATCCACAGC CAAAGACACT 480
 CTAGACAATG GCGATTATGG AGAACGAGC TATGACTGGA CCACGGGCCC CAGGGACGAC 540
 40 GACGAGTCTG ATNGACACCT TGAAGAAAA CAGGGGTAC ATGGAAATG AACAGTCAGT 600
 GAAATCTTTT AAGATGCCAT CCTAAATAT AGAAGAGGAA GACAGCCATT TCTTTTCTCA 660
 45 TCTTATTATT TTCTCTTTT GAATTCGCTG TGTTACATT ACATATCACA AAAAAAGGAA 720
 GATTTTCTTT CTGTTTCAAA GCAAGAAATG GCGTGATGTC CTTTGTCCCA AAACAGTGG 780
 ATACCATGCC CTAGATCAGA ATGTTAATGA GGCAATGCTT TCTTTGAAGA TTACCAATGA 840
 50 TTATAATTTT TAAAGCACTG TCATTGAAT TIGCTTATGT AATTATATT GTTGAATTT 900
 TTTTATTTT TAAAGCACTG TCATTGAAT TIGCTTATGT AATTATATT GTTGAATTT 960
 TTTTATTTT TAAAGCACTG TCATTGAAT TIGCTTATGT AATTATATT GTTGAATTT 1020
 TTTTATTTT TAAAGCACTG TCATTGAAT TIGCTTATGT AATTATATT GTTGAATTT 1080
 TTTTATTTT TAAAGCACTG TCATTGAAT TIGCTTATGT AATTATATT GTTGAATTT 1140
 60

TGCTACTTTT AAAAGATCCC AAACCTGTAA CTAAATTCTG ACATATCTGT TACTGCTGAC 1200
 TDACATTCAT TCTCCGCCAT TCAAATACFA TTTTMTATCC ACATTTTTTT TTGTTCCCAA 1260
 5 ACTGTAATGT ACAAGGATAT GTGTGATAAT GCTTTGGATT TGAGTAATAT TTTTTTTTCT 1320
 TCCAAGAAAA CTGCTTTGGA TATTTTGA TAATTTAAAC ATAATTTAGG ATAATGATAT 1380
 10 TGCTCAATCT GACCACAATT TTAGGTAAAA CATTAAATGT GTCAAGAAAT CTGGCAACA 1440
 GAACTCTCTC AGCTTGCAST GGACATAGAT AAAATGTTAC AGAGATACTA TTTTTTGGT 1500
 TGAATTACT ATATTAAAT TAGAAGCAGA AACTGGTAAA ATGTTAAATA CATGTACAAT 1560
 15 TGTMTTAST TAGCAATGA TTGTAGCATG GGTTCCTCCA AGGTTTCAAG CAATGGGCAG 1620
 AGTTTAAAT TATATCAGAT TCGTTACTT GGTATTAT TTTACAGTAA ATTTGAATAA 1680
 ATCTTAGGG TCATTATCAC TTAAATAATA CTCTACCTAG GTCTTTCAA TTAATAATTAT 1740
 20 AACTGAATGA AGTTGTTGT ATACATAAAG GATATTTGTG TACAATTACC TTTTTCCC 1800
 CAACTTGT TTCTTTGTT TTGTTTTTA TGCCAACCTG AAAGTATTTA CTATGGGATT 1860
 25 CATTATGTC TGCTTTTA TCATAAAGAA TGAATCAATA TGTAATATG TGATTGAAC 1920
 CATGGTTGAC TTACAAGTGT CACTACAGCT TTTTAGAAAA CATAGCCCTA ATATATGTTA 1980
 AGCAGGACCC GGGTAGGCA GTGGGCTTGC GCTTTATGTA GAGCTGAAG AAGCCCTCC 2040
 30 ATCCTGTCTC TTGGGCGGAC AGTGTACTTT CCTAATAGGG AAGGGAAGCA CAATGGAAAT 2100
 ACCCTGAAC CGTTTTATTG CAGTAATTTT TTTCATATCT GAACTATTA TTTAATATTT 2160
 35 TGAATAAGAT TTTAAAAAT AAATGGCAAA GATATAAATC TAAAAAANA AAAAAAANA 2220
 AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA N 2271

40

(2) INFORMATION FOR SEQ ID NO: 233:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

55

60

CTCCGGTTC TCCGGGCAGC TCCACTGCT GTAGCTTCTG CCACCTGCCA CGACCGGGCC 60
 TCTCCCTGGC GTTTGCTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCCTCTTGA 120
 GCCAGCGTGS CGNGCCTGGC GGCTCCCGGG TGGTGAGAGA GCGGTCCGGG AACATGAAG 180
 GCCTGGCAGT GCTGCTGCTG TCTCAGGCAC CTCTTGGCTT CCGTCCCTCT CTTGCTGTTG 240
 CTGCTGAAC TAAGCGGTC CTTGGMASTC CTGCTGCAGG CAGCCGAGGC CGGCCCAGGT 300

YTTGGGGCTC CTGACCCTAG ACCAGGACAT TACCGCCGCT GCCACGGGGC CCTWACCCCT 360
 GCGCAGCAGC CCGGCTCTGG TGTGGTCAA GCTGGGGGGG CCGCGGGGCT CCGAGGGAGG 420
 5 CAATGGCAGC AACCTGTGG CCGGCTTGA GACGGACGAT CACGGACCGA AGCGCGGGA
 ARGCTGGTG GGTGCGGGT TGTCTGAG CCGCAACCT GCGACAAGC CCATGACCA 540
 10 GCGGCGCTG ACCGTGTTGA TGSTGGTGA CCGCGGGTG CTGCTGACT TCGTGGTCAG 600
 GACGTCAGC ATGAGAACAA GAAACCGAAA GACTAGGAGA TATGGAGTIT TCGACACTAA 660
 CATAGAAAAT ATGGAATTGA CACCTTTAGA ACAGGATGAT GAGGATGATG ACAACACGTT 720
 15 GTTTGATGCC AATCATCTC GAAGATAAGA ATGTGCCCTT TGATGAAAGA ACTTTATCTT 780
 TCTACAATGA AGAGTGGAAT TTCTATGTTT AAGGAATAAG AAGCCACTAT ATCAATGTTG 840
 20 GGGGGGTATT TAAGTTACAT ATATTTAAAC AACCTTTAAT TTGCTGTTC AATAAATACC 900
 GTATCTTTT ATTATATCTT TATATGTATA GAGTACTCT GTTAATGGGC TCAGAGATGT 960
 TGGGGATAAA GTATACTGTA ATAATTTATC TTTTIGAAAA TTAATAAAA ACGGTGTTTT 1020
 25 CTGTGGGTT TTTGTTTCT GCTTACATA TGATTGTAAA TTGTTTATG TATTAATCAG 1080
 TTAATGCTAA TTATTTTTC TGATGTATA TGTTAAAGAG CTATAAATTC CAACAACCAA 1140
 30 CTGGTGTGA AAAATAATTT AAAATYTCTT TTAATGAAAG GTATTTCCCA TTTTGTGGG 1200
 GAAAAGAAGC CAAATTTATT ACTTTGTGTT GGGGTTTTTA AATATTAAG AATGTCTAA 1260
 GTTATTGTTT GCAAAACAAT AAATATGATT TAAATTCTC TAAAAAAA AAAAAAAC 1320
 35 CCGGGGGGG GCGCCGNN 1338

40

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

50 Met Leu Ser Thr Gly Ile Glu Val Ala Arg Pro Pro Ala Thr Leu Leu
 1 5 10 15
 Gly Leu Met Phe Val Leu Thr Gly Met Pro Arg Gly Leu Arg Xaa
 20 25 30

60 (i) SEQUENCE CHARACTERISTICS
 (A) LENGTH: 119 amino acids

488

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

5 Met Asn Val Val Ile Val Ile Ile Leu Phe Ser Phe Asp Ser Val Gly
 1 5 10 15

Thr Met Phe Ser Cys Asn Arg Ile Pro Lys Ile Thr Val Leu Asn Lys
 20 25 30

10 Leu Lys Phe Xaa Cys Glu Val Leu Leu Arg Ile Gln Thr Ile Gln Gly
 35 40 45

15 Phe Tyr Arg Cys Thr Arg Ile Ser Arg Tyr Lys Gly Ile Phe Pro Asp
 50 55 60

Phe Cys Gln Ser Gln Cys Met Gly Cys Asn Pro Glu Ser Xaa Met Ala
 65 70 75 80

20 Val Pro Ala Leu Val Thr Pro Ile Leu Ala His Arg Lys Lys Glu Lys
 85 90 95

Gly Met Cys Leu Phe Thr Leu Ile Ile Ala Pro Thr Arg Cys Thr His
 100 105 110

25 Tyr Phe Cys Xaa
 115

30

(2) INFORMATION FOR SEQ ID NO: 236:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

40 Met Ser Ser Ala Lys Ile Val Arg Gln Arg Gly Ala Val Pro Thr Tyr
 1 5 10 15

Tyr Thr Thr Glu Ala Gly Glu Ile Ile Phe Leu Val Leu Asn Trp Ser
 20 25 30

45 Leu Ser Ile Leu His Ile Val Asp Val Leu Cys Ser Lys Pro Glu Lys
 35 40 45

Ser Val Thr Glu Asp Ala Ala Ser Gly Leu Ser Gln Arg Met Thr Ala
 50 55 60

50 Leu Val Trp Arg Lys Gly Pro Asp Gly Gly Ser Arg Lys Pro Ile Leu
 65 70 75 80

55 Leu Leu Phe Phe Phe Leu Pro Leu Ile Leu Cys Phe His Ser Phe Ile
 85 90 95

His Ser Ser Asn Ile Cys Xaa
 100

60

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

10 Met Ile Leu Phe Pro Gln Xaa Ala Leu Arg Leu Gly Xaa Trp Pro Arg
 1 5 10 15
 Thr Trp Ser Ile Leu Xaa Lys Tyr Ser Val Asn Phe Phe Ser Ala Tyr
 20 25 30
 15 Ser Pro Met Gly Ala Val Gly Thr Glu Phe
 35 40

20

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 37 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

30 Met Ile Ile Leu Leu Leu Phe Met Leu Leu Asn Asn Val Val Leu Val
 1 5 10 15
 Gln Glu Asp Asn Cys Gln Arg Lys Asn Thr Val Gln Glu Arg Arg Xaa
 20 25 30
 35 Trp Ser Gln Trp Xaa
 35

40 (2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 128 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

50 Met Ala Ala Xaa Pro Pro Gly Cys Thr Pro Pro Xaa Leu Leu Asp Ile
 1 5 10 15
 Ser Trp Leu Thr Glu Ser Leu Gly Ala Gly Gln Pro Val Pro Val Glu
 20 25 30

60 His His Arg Thr Gly Leu Ser Glu Leu Leu Glu His Gly Val Cys Gln

65					70						75					80
Glu	Val	Glu	Arg	Val	Arg	Arg	Ser	Glu	Arg	Tyr	Gln	Thr	Met	Lys	Val	
				85					90					95		
Arg	Arg	Ala	Gly	Leu	Gly	Pro	Thr	Pro	Gly	Met	Ser	Cys	Pro	Gly	Asn	
			100					105					110			
Asp	Asn	Thr	Val	His	Thr	Met	His	Gly	Glu	Ala	Asn	Arg	Gly	Ser	Xaa	
	115						120					125				

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

```

Met Ser Ile Leu Cys Cys Pro Xaa Leu Cys Leu Phe Phe Ser Phe Cys
  1                      5                      10                      15

Ile Ser Ser Gly Ser Cys Pro Phe Ser His Val Ser Gln Leu Ser Phe
      20                      25                      30

Ile Ala Thr Phe Ser Gln Ser Ser Pro Val Leu Leu Val Pro Ala Tyr
      35                      40                      45

Asn Thr Tyr Leu Ser Phe Leu Ala Phe Leu Asp Cys Ala Ser Leu Thr
      50                      55                      60

Ser Thr Xaa
      65

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Met Ser Thr Phe Gln Leu Leu Leu Ile Leu Ala Gln Ser Thr Tyr
1 5 10 15
Lys Ile Lys Ser Lys Pro Leu His Met Thr Asn His Thr Leu Leu Asn
20 25 30
Ser Pro Gly Leu Asn Pro Ser Ser Pro Thr Leu Asn Phe Lys Thr Gln
35 40 45
Gln His Glu Ser Val Ser Tyr Ala Cys Cys His Met Arg Ser Leu His
50 55 60

491

His Ala Phe Ala Xaa
65

5

(2) INFORMATION FOR SEQ ID NO: 242:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 44 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

15 Met Val Ser Val Val Leu Ile Phe Ser Phe Leu Ser Leu Thr Ile Ser
 1 5 10 15
 Thr Thr Ala Ser Ala Tyr Asn Gly Asn Asp Thr Gln Gly Trp Asn Asp
 20 25 30
 Lys Phe His Xaa Xaa Ser Val Lys Thr Gln Thr Xaa
 35 40

25

(2) INFORMATION FOR SEQ ID NO: 243:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

35 Met Ile Ser Asp Ala Gly Ala Gly Phe Gly Val Phe Leu Leu Val Pro
 1 5 10 15
 Arg Ala Gly His Cys Trp Gly Ala Gly Lys Pro Leu Pro Ser Cys Pro
 20 25 30
 40 Ser Val Ala Ser Ile Pro Ser Trp Val Leu Pro Ser Phe Leu Glu Arg
 35 40 45
 Gly Arg Xaa
 50

45

(2) INFORMATION FOR SEQ ID NO: 244:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 43 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

60

Phe Leu Leu Leu Leu Met Phe Gln Thr Leu Ser Leu Ala Pro Ala Thr
 20 25 30

Leu Lys Pro Leu Arg Val Thr Ser His Ser Xaa
35 40

5

(2) INFORMATION FOR SEQ ID NO: 245:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 61 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

15 Met Ile Leu Met Pro Gly Leu Gly Thr Ser Arg Gln Arg Ser Val Pro
1 5 10 15
 Phe Val Pro Thr Leu Asn Ala Ser Thr Pro Gly Ala Met Thr Gly Pro
20 25 30
 Thr Ala Thr Leu Thr Ser Cys Gln Trp Thr Thr Ala Cys Arg Val Ser
35 40 45
 Trp Ala Asn Gly Trp Thr Ser Leu Arg Thr Phe Arg Xaa
25 50 55 60

30

(2) INFORMATION FOR SEQ ID NO: 246:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Ser His His Ala Gln Pro Arg Phe Leu Leu Ile Thr Met Leu Leu
1 5 10 15
 Gln Glu Ala Lys Pro Val Ser Asn Ile Pro His Leu Leu Glu Ser Trp
20 25 30
 Tyr Phe Gly Xaa
35

45

(2) INFORMATION FOR SEQ ID NO: 247:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 38 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55

Met Asn Ser Leu Phe Trp Met Ile Leu Leu Pro Val Ser Gln Asp Gln
1 5 10 15
 Val Val Glu Gly Leu Gln Gly Gly Phe Ser Gln Ile His Met Arg Ile
20 25 30

60

493

Leu Arg Lys His Leu Xaa
35

5

(2) INFORMATION FOR SEQ ID NO: 248:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 211 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

15

Met Ser Arg Ser Xaa Asp Val Thr Asn Thr Thr Phe Leu Leu Met Ala
1 5 10 15

Ala Ser Ile Tyr Leu His Asp Gln Asn Pro Asp Ala Ala Leu Arg Ala
20 25 30

20

Leu His Gln Gly Asp Ser Leu Glu Cys Thr Ala Met Thr Val Gln Ile
35 40 45

25

Leu Leu Lys Leu Asp Arg Leu Asp Leu Ala Arg Lys Glu Leu Lys Arg
50 55 60

Met Gln Asp Leu Asp Glu Asp Ala Thr Leu Thr Gln Leu Ala Thr Ala
65 70 75 80

30

Trp Val Ser Leu Ala Thr Gly Gly Glu Lys Leu Gln Asp Ala Tyr Tyr
85 90 95

Ile Phe Gln Glu Met Ala Asp Lys Cys Ser Pro Thr Leu Leu Leu Leu
100 105 110

35

Asn Gly Gln Ala Ala Cys His Met Ala Gln Gly Arg Trp Glu Ala Ala
115 120 125

40

Glu Gly Leu Leu Gln Glu Ala Leu Asp Lys Asp Ser Gly Tyr Pro Glu
130 135 140

Thr Leu Val Asn Leu Ile Val Leu Ser Gln His Leu Gly Lys Pro Pro
145 150 155 160

45

Glu Val Thr Asn Arg Tyr Leu Ser Gln Leu Lys Asp Ala His Arg Ser
165 170 175

His Pro Phe Ile Lys Glu Tyr Gln Ala Lys Glu Asn Asp Phe Asp Arg
180 185 190

50

Leu Val Leu Gln Tyr Ala Pro Ser Ala Glu Ala Gly Pro Glu Leu Ser
195 200 205

60

(2) INFORMATION FOR SEQ ID NO: 249:

494

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 548 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Met Glu Asp Ser Glu Ala Leu Gly Phe Glu His Met Gly Leu Asp Pro
 1 5 10 15

10 Arg Leu Leu Gln Ala Val Thr Asp Leu Gly Trp Ser Arg Pro Thr Leu
 20 25 30

Ile Gln Glu Lys Ala Ile Pro Leu Ala Leu Glu Gly Lys Asp Leu Leu
 35 40 45

15 Ala Arg Ala Arg Thr Gly Ser Gly Lys Thr Ala Ala Tyr Ala Ile Pro
 50 55 60

20 Met Leu Gln Leu Leu Leu His Arg Lys Ala Thr Gly Pro Val Val Glu
 65 70 75 80

Gln Ala Val Arg Gly Leu Val Leu Val Pro Thr Lys Glu Leu Ala Arg
 85 90 95

25 Gln Ala Gln Ser Met Ile Gln Gln Leu Ala Thr Tyr Cys Ala Arg Asp
 100 105 110

Val Arg Val Ala Asn Val Ser Ala Ala Glu Asp Ser Val Ser Gln Arg
 115 120 125

30 Ala Val Leu Met Glu Lys Pro Asp Val Val Val Gly Thr Pro Ser Arg
 130 135 140

Ile Leu Ser His Leu Gln Gln Asp Ser Leu Lys Leu Arg Asp Ser Leu
 145 150 155 160

Glu Leu Leu Val Val Asp Glu Ala Asp Leu Leu Phe Ser Phe Gly Phe
 165 170 175

40 Glu Glu Glu Leu Lys Ser Leu Leu Cys His Leu Pro Arg Ile Tyr Gln
 180 185 190

Ala Phe Leu Met Ser Ala Thr Phe Asn Glu Asp Val Gln Ala Leu Lys
 195 200 205

45 Glu Leu Ile Leu His Asn Pro Val Thr Leu Lys Leu Gln Glu Ser Gln
 210 215 220

Leu Pro Gly Pro Asp Gln Leu Gln Gln Phe Gln Val Val Cys Glu Thr
 225 230 235 240

Glu Glu Asp Lys Phe Leu Leu Leu Tyr Ala Leu Leu Lys Leu Ser Leu
 245 250 255

50 Ile Arg Gly Lys Ser Leu Leu Phe Val Asn Thr Leu Glu Arg Ser Tyr
 260 265 270

Arg Leu Arg Leu Phe Leu Glu Gln Phe Ser Ile Pro Thr Cys Val Leu
 275 280 285

60

495

Asn Gly Glu Leu Pro Leu Arg Ser Arg Cys His Ile Ile Ser Gln Phe
 290 295 300

5 Asn Gln Gly Phe Tyr Asp Cys Val Ile Ala Thr Asp Ala Glu Val Leu
 305 310 315 320

Gly Ala Pro Val Lys Gly Lys Arg Arg Gly Arg Gly Pro Lys Gly Asp
 325 330 335

10 Lys Ala Ser Asp Pro Glu Ala Gly Val Ala Arg Gly Ile Asp Phe His
 340 345 350

His Val Ser Ala Val Leu Asn Phe Asp Leu Pro Pro Thr Pro Glu Ala
 355 360 365

15 Tyr Ile His Arg Ala Gly Arg Thr Ala Arg Ala Asn Asn Pro Gly Ile
 370 375 380

Val Leu Thr Phe Val Leu Pro Thr Glu Gln Phe His Leu Gly Lys Ile
 385 390 395 400

Glu Glu Leu Leu Ser Gly Glu Asn Arg Gly Pro Ile Leu Leu Pro Tyr
 405 410 415

25 Gln Phe Arg Met Glu Glu Ile Glu Gly Phe Arg Tyr Arg Cys Arg Asp
 420 425 430

Ala Met Arg Ser Val Thr Lys Gln Ala Ile Arg Glu Ala Arg Leu Lys
 435 440 445

30 Glu Ile Lys Glu Glu Leu Leu His Ser Glu Lys Leu Lys Thr Tyr Phe
 450 455 460

Glu Asp Asn Pro Arg Asp Leu Gln Leu Leu Arg His Asp Leu Pro Leu
 465 470 475 480

His Pro Ala Val Val Lys Pro His Leu Gly His Val Pro Asp Tyr Leu
 485 490 495

40 Val Pro Pro Ala Leu Arg Gly Leu Val Arg Pro His Lys Lys Arg Lys
 500 505 510

Lys Leu Ser Ser Ser Cys Arg Lys Ala Lys Arg Ala Lys Ser Gln Asn
 515 520 525

45 Pro Leu Arg Ser Phe Lys His Lys Gly Lys Lys Phe Arg Pro Thr Ala
 530 535 540

50 Lys Pro Ser Xaa
 545

60

(i) SEQUENCE DESCRIPTION: SEQ ID NO. 250:

Met Thr Thr Val Pro Pro Ser Pro Arg Pro Met Ser Arg Pro Ser Glu
 1 5 10 15
 5 Arg Asn Met Arg Arg Pro Arg Gly Pro Ser Pro Leu Pro Ala Ser Pro
 20 25 30
 Arg Asn Ser Thr Pro Asp Glu Pro Asp Val His Phe Ser Lys Lys Phe
 35 40 45
 10 Leu Asn Val Phe Met Ser Gly Arg Ser Arg Ser Ser Ser Ala Glu Ser
 50 55 60
 15 Phe Gly Leu Phe Ser Cys Ile Ile Asn Gly Glu Glu Gln Glu Gln Thr
 65 70 75 80
 His Arg Ala Ile Phe Arg Phe Val Pro Arg His Glu Asp Glu Leu Glu
 85 90 95
 20 Leu Glu Val Asp Asp Pro Leu Leu Val Glu Leu Gln Ala Glu Asp Tyr
 100 105 110
 Trp Tyr Glu Ala Tyr Asn Met Arg Thr Gly Ala Arg Gly Val Phe Pro
 115 120 125
 25 Ala Tyr Tyr Ala Ile Glu Val Thr Lys Glu Pro Glu His Met Ala Ala
 130 135 140
 30 Leu Ala Lys Asn Ser Asp Trp Val Asp Gln Phe Arg Val Lys Phe Leu
 145 150 155 160
 Gly Ser Val Gln Val Pro Tyr His Lys Gly Asn Asp Val Leu Cys Ala
 165 170 175
 35 Ala Met Gln Lys Ile Ala Thr Thr Arg Arg Leu Thr Val His Phe Asn
 180 185 190
 Pro Pro Ser Ser Cys Val Leu Glu Ile Ser Val Arg Gly Val Lys Ile
 195 200 205
 40 Gly Val Lys Ala Asp Asp Ser Gln Glu Ala Lys Gly Asn Lys Cys Ser
 210 215 220
 45 His Phe Phe Gln Leu Lys Asn Ile Ser Phe Cys Gly Tyr His Pro Lys
 225 230 235 240
 Asn Asn Lys Tyr Phe Gly Phe Ile Thr Lys His Pro Ala Asp His Arg
 245 250 255
 50 Phe Ala Cys His Val Phe Val Ser Glu Asp Ser Thr Lys Ala Leu Ala
 260 265 270
 Glu Ser Val Gly Arg Ala Phe Gln Gln Phe Tyr Lys Gln Phe Val Glu
 275 280 285
 55 Tyr Thr Cys Pro Thr Glu Asp Ile Tyr Leu Glu
 290 295
 60

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

5
10 Leu Leu Tyr Leu Leu Lys Val Xaa Val Ile Phe Val Phe Ser Ser Ser
1 5 10 15
Lys Gly Val Thr Leu Val Ser Met Asn Leu Thr Ser Phe Phe Val Ser
20 25 30
15 Ser Val Leu Ala Cys Phe Ser Xaa
35 40

20 (2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 594 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

25
30 Met Pro Ala Ser Ser Leu Glu Ser Arg Ser Phe Leu Leu Ala Lys Lys
1 5 10 15
Ser Gly Glu Asn Val Ala Lys Phe Ile Ile Asn Ser Tyr Pro Lys Tyr
20 25 30
35 Phe Gln Lys Asp Ile Ala Glu Pro His Ile Pro Cys Leu Met Pro Glu
35 40 45
Tyr Phe Glu Pro Gln Ile Lys Asp Ile Ser Glu Ala Ala Leu Lys Glu
50 55 60
40 Arg Ile Glu Leu Arg Lys Val Lys Ala Ser Val Asp Met Phe Asp Gln
65 70 75 80
Leu Leu Gln Ala Gly Thr Thr Val Ser Leu Glu Thr Thr Asn Ser Leu
85 90 95
45 Leu Asp Xaa Leu Cys Tyr Tyr Gly Asp Gln Glu Pro Ser Thr Asp Tyr
100 105 110
50 His Phe Gln Gln Thr Gly Gln Ser Glu Ala Leu Glu Glu Glu Asn Asp
115 120 125
Glu Thr Ser Arg Arg Lys Ala Gly His Gln Phe Gly Val Thr Trp Arg
130 135 140

60
145 150 155

498

Tyr Glu Gln Ala Leu Asn Leu Tyr Thr Glu Leu Leu Asn Asn Arg Leu
 180 185 190

5 His Ala Asp Val Tyr Thr Phe Asn Ala Leu Ile Glu Ala Thr Val Cys
 195 200 205

Ala Ile Asn Glu Lys Phe Glu Glu Lys Trp Ser Lys Ile Leu Glu Leu
 210 215 220

10 Leu Arg His Met Val Ala Gln Lys Val Lys Pro Asn Leu Gln Thr Phe
 225 230 235 240

Asn Thr Ile Leu Lys Cys Leu Arg Arg Phe His Val Phe Ala Arg Ser
 245 250 255

15 Pro Ala Leu Gln Val Leu Arg Glu Met Lys Ala Ile Gly Ile Glu Pro
 260 265 270

20 Ser Leu Ala Thr Tyr His His Ile Ile Arg Leu Phe Asp Gln Pro Gly
 275 280 285

Asp Pro Leu Lys Arg Ser Ser Phe Ile Ile Tyr Asp Ile Met Asn Glu
 290 295 300

25 Leu Met Gly Lys Arg Phe Ser Pro Lys Asp Pro Asp Asp Asp Lys Phe
 305 310 315 320

Phe Gln Ser Ala Met Ser Ile Cys Ser Ser Leu Arg Asp Leu Glu Leu
 325 330 335

30 Ala Tyr Gln Val His Gly Leu Leu Lys Thr Gly Asp Asn Trp Lys Phe
 340 345 350

35 Ile Gly Pro Asp Gln His Arg Asn Phe Tyr Tyr Ser Lys Phe Phe Asp
 355 360 365

Leu Ile Cys Leu Met Glu Gln Ile Asp Val Thr Leu Lys Trp Tyr Glu
 370 375 380

40 Asp Leu Ile Pro Ser Ala Tyr Phe Pro His Ser Gln Thr Met Ile His
 385 390 395 400

Leu Leu Gln Ala Leu Asp Val Ala Asn Arg Leu Glu Val Ile Pro Lys
 405 410 415

45 Ile Trp Lys Asp Ser Lys Glu Tyr Gly His Thr Phe Arg Ser Asp Leu
 420 425 430

Arg Glu Glu Ile Leu Met Leu Met Ala Arg Asp Lys His Pro Pro Glu
 435 440 445

Leu Gln Val Ala Phe Ala Asp Cys Ala Ala Asp Ile Lys Ser Ala Tyr
 450 455 460

55 Glu Ser Gln Pro Ile Arg Gln Thr Ala Gln Asp Trp Pro Ala Thr Ser
 465 470 475 480

Leu Asn Cys Ile Ala Ile Leu Phe Leu Arg Ala Gly Arg Thr Gln Glu
 485 490 495

60